LISTENER ANONYMIZER: CAMOUFLAGING PLAY LOGS TO PRESERVE USER'S DEMOGRAPHIC ANONYMITY

Kosetsu Tsukuda **Satoru Fukayama** **Masataka Goto**
National Institute of Advanced Industrial Science and Technology (AIST), Japan
{k.tsukuda, s.fukayama, m.goto}@aist.go.jp

ABSTRACT

When a user signs up with an online music service, she is often requested to register her demographic attributes such as age, gender, and nationality. Even if she does not input such information, it has been reported that user attributes can be predicted with high accuracy by using her play log. How can users enjoy music when using an online music service while preserving their demographic anonymity? To solve this problem, we propose a system called *Listener Anonymizer*. Listener Anonymizer monitors the user's play log. When it detects that her confidential attributes can be predicted, it selects songs that can decrease the prediction accuracy and recommends them to her. The user can camouflage her play logs by playing these songs to preserve her demographic anonymity. Since such songs do not always match her music taste, selecting as few songs as possible that can effectively anonymize her attributes is required. Listener Anonymizer realizes this by selecting songs based on feature ablation analysis. Our experimental results using Last.fm play logs showed that Listener Anonymizer was able to preserve anonymity with fewer songs than a method that randomly selected songs.

1. INTRODUCTION

When a user signs up with an online music service (*e.g.*, Last.fm [1] and Spotify [2]), it is common for the user to be asked to input her demographic attributes such as age and gender. Registering such demographic attributes is beneficial for her because various songs are recommended to her by the service according to her attributes. In addition, she can follow another user who has similar demographic attributes, and they can communicate with each other. Despite such benefits, many users conceal their demographic attributes because they would be concerned about privacy. As shown in Section 5.1, as many as 49.3% of Last.fm users do not register any of the age, gender, and nationality attributes. If a user does not register her demographic attributes, is her privacy fully protected?

Several studies have aimed to predict users' demographic attributes from their music play logs [10, 12, 25]. They have tackled the problem because it had been reported that the attributes contribute to improving music recommendation accuracy [21, 23, 27]. However, users who do not register their attributes might not want researchers or companies to predict their demographic attributes. It is not only a psychological problem; if a user's demographic attributes are predicted, she may suffer damage. For example, suppose one day the email address and music play logs of a female user who has not input her gender on an online music service are leaked from the website, and a malicious company obtains the data. If the malicious company can predict her gender with high accuracy from the logs, it can send her spam e-mails that target females.

What should we do to enable users to enjoy music when using an online music service while preserving their demographic anonymity? In this paper, we propose a system called *Listener Anonymizer* to solve this problem. Listener Anonymizer camouflages the user's play log and preserves the anonymity of her confidential attributes. To be more specific, when a user plays a song using an online music service, Listener Anonymizer monitors the songs that are played. If Listener Anonymizer detects that the user's confidential attributes can be predicted with an accuracy above a certain level, the system selects songs that can decrease the prediction accuracy and recommends them to her. The user can camouflage her play log by playing them and preserve her anonymity. However, since such selected songs do not always match her music taste, selecting as few songs as possible that can effectively anonymize her attributes is required. To achieve this, we propose a method for selecting songs according to the user's confidential attributes.

Our contributions in this paper are as follows.

- To the best of our knowledge, this is the first study that introduces the concept of preserving the anonymity of the users' demographic attributes while they play songs using an online music service.
- We propose an approach that camouflages the user's play log to preserve her anonymity. We marshaled factors to consider for selecting songs from five viewpoints: the definition of anonymity, method for predicting demographics, timing of camouflaging play logs, user's true demographics, and anonymization of multiple demographics.
- To examine the effectiveness of the proposed method, we carried out experiments using Last.fm play logs. Our experimental results showed that our proposed method was able to preserve anonymity with fewer songs than a method that randomly selected songs. Based on the experiments, we dis-

[1] http://www.last.fm
[2] http://www.spotify.com

cuss four important considerations: impact on recommendation accuracy, user's taste in music, simulation of multiple prediction methods, and real-time monitoring of songs that are played.

2. RELATED WORK

Since predicting the user's demographic attributes can be used in many applications such as content recommendation and user behavior analysis, studies of demographic prediction have been conducted in various domains. One of the most popular domains is social media such as Twitter [3] and Facebook [4]. It is known that language use on social media varies according to demographic attributes such as age [9] and gender [7]. Hence, most studies have used text data posted to social media and utilized machine learning techniques to predict users' demographic attributes [15,18,24]. Although it was thought that predicting demographics was a difficult task [15, 16], recent studies reported a high prediction accuracy. For example, age and gender can be predicted with mean absolute error (MAE) of 3.40 and a binary classification accuracy of 91.9%, respectively [18]. In addition to social media, demographic prediction has been conducted in the fields of blogs [2, 5] and web search queries [8].

In the field of music information retrieval (MIR), too, users' demographic attributes on an online music service play an important role mainly for music recommendation. As reported by Uitdenbogerd and Schnydel [22], music preference is affected by individual factors including age and ethnicity. In fact, it was revealed that music recommendation accuracy was improved by considering demographic attributes [21, 23, 27]. Motivated by these results, several studies in MIR have aimed to predict demographic attributes. The main way to perform this task is to use play logs obtained from an online music service and supervised machine learning techniques. Liu and Yang [12] predicted age and gender by using timestamps, song/artist metadata, and acoustic features of music signals. Wu *et al.* [25] also proposed methods to predict age and gender based on music metadata. They created two kinds of features: a TF-IDF-based one and a GSV(Gaussian super vector [3])-based one. They applied support vector machine (SVM) to them. Krismayer *et al.* [10] predicted nationality in addition to age and gender based on music metadata (artist names and artist's tags). The details of their method are described in Section 4.2. By using their method, it was reported that demographic attributes can be predicted with high accuracy. The age was predicted with MAE of 4.13, and the gender and nationality were predicted with a classification accuracy of 81.36% and 69.37%, respectively.

Unlike these studies, our goal is to preserve users' demographic anonymity since some users do not want researchers or companies to predict their demographic attributes. Although several studies have discussed privacy problems (*e.g.*, the release of a user query log can lead to loss of privacy [8], confidential information such as medical conditions can be inferred from tweets [14], and how should researchers deal with personal information in MIR [19]), our study is different from these studies in that we propose a concrete anonymization system and carried out experiments to evaluate how well it works.

[3] https://twitter.com
[4] https://facebook.com

3. FACTORS FOR REALIZING LISTENER ANONYMIZER

As we described in Section 1, we propose an approach that selects songs and camouflages the user's play log by playing these songs so that the user can preserve demographic anonymity. To enable an intuitive understanding of our idea, we give the following example story.

Emma is a 22-year-old French female. She is a Last.fm user and concealed her nationality when she signed up. She also uses Listener Anonymizer, which monitors the music she plays using Last.fm. One day, when Emma is listening to music using Last.fm with her smartphone, Listener Anonymizer detects that her nationality can be predicted as French with high accuracy from her play logs. Thus, Listener Anonymizer shows an alert message stating "your nationality can be predicted as French with a probability of 67%" on her smartphone screen and recommends three songs to her. Emma plays the songs to preserve the anonymity of her nationality.

Although this is just an example story, we need to consider various factors to realize Listener Anonymizer. Below, we marshal the factors from five viewpoints.

3.1 Definition of Anonymity

First, we define the anonymity of demographic attributes. In this paper, we propose two kinds of concepts for anonymity: *not-first-anonymity* and *k-flat-anonymity*. Suppose a demographic attribute d has n attribute values represented by $A_d = \{a_1, a_2, \cdots, a_n\}$. For example, when d is nationality, $a_i \in A_d$ can be French, Japanese, etc. When user u has an attribute value $a_u \in A_d$ and conceals the attribute, given her music play log, we can compute the probability $p(a_i)$ for each attribute value in A_d by using an attribute prediction method ($0 \leq p(a_i) \leq 1$ and $\sum_{i=1}^{n} p(a_i) = 1$). In this case, not-first-anonymity is satisfied if the following condition is met: the rank of $p(a_u)$ among all attribute values is not the highest. In the case of Emma, not-first-anonymity is satisfied when the probability of French is not the highest.

In the case of k-flat-anonymity, the anonymity is satisfied if the following condition is met. Given the top k attribute values in terms of the probability, a_u is included in the top k attribute values and the probability gap between any two attribute values is lower than θ. In k-flat-anonymity, user's demographic attributes may be regarded as unpredictable because the top k attribute values have almost the same probabilities. In the example of Emma, suppose k and θ are set to 3 and 0.05, respectively, and the probabilities of French, Spanish, and German are 0.32, 0.28, and 0.29, respectively. In this case, because the probabilities of other nationalities are lower than those of the three nationalities and the probability gap between any two nationalities out of the three nationalities is lower than 0.05, k-flat-anonymity is satisfied even though French has the highest probability.

3.2 Method for Predicting Demographics

To realize Listener Anonymizer, simulating the method used in a demographics prediction system is required so that we can show an alert at the right time. However, since it is generally impossible to know the prediction method, we have to assume some prediction methods and propose a song selection method according to them. It is common to use music metadata extracted from the user's play log for predicting demographic attributes [10, 12, 25]. If we propose a song selection method that works well for state-of-the-art methods that are based on music metadata, we can say that our proposed method is robust to a certain extent. In light of the above, in this paper, we propose a method for selecting songs in Section 4.3 and show the effectiveness of this method through experiments in Section 5.

3.3 Timing of Camouflaging Play Logs

Listener Anonymizer selects songs and camouflages play logs in two main situations. One is when not-first-anonymity (or k-flat-anonymity) is no longer satisfied as we described in the example story at the beginning of this section. The other is when a user does not use a smartphone such as when she is sleeping or taking a bath. In the former case, since the user listens to her favorite songs before the songs are recommended by Listener Anonymizer, it should select as few songs as possible so that the user can soon resume listening to her favorite songs. In the latter case, since the user has enough time to play recommended songs and does not need to listen to them, a method that randomly selects many songs might be enough for recovering not-first-anonymity. Some users still hope to play as few recommended songs as possible to save on the packet communication fee.

3.4 User's True Demographics

In the preceding sections, we assumed that our anonymization system knows the user's true attribute values (*e.g.*, Emma's nationality is French). That is, the user has to input the true demographic attributes before starting to use Listener Anonymizer. However, some users would not want to tell even the system their true demographics. When the system does not know the user's true demographic attribute, not-first-anonymity can be defined as follows: when the difference between the highest probability and the second highest probability is lower than θ. In this case, not-first-anonymity will not often be satisfied, and songs will be more frequently recommended to the user than the case where the system knows the true demographic attribute. In the example of Emma, suppose she does not tell her nationality to Listener Anonymizer. If she wants to preserve complete anonymity, she must play recommended songs at every alert, but this is a heavy burden for her. She could ignore an alert if the predicted nationality is wrong. However, if she plays only the recommended songs when the predicted nationality is French, Listener Anonymizer can estimate that Emma's nationality is French.

When the anonymization system knows a user's true attribute, the alert is displayed only when the true demographic can be predicted, which reduces the user's burden. In addition, if we can implement Listener Anonymizer as a stand-alone smartphone application, the user's true demographics are stored only in the smartphone and are not sent to a server. In this case, users do not need to worry about leakage of demographic information from the server.

3.5 Anonymization of Multiple Demographics

We need to consider a situation where a user wants to preserve anonymity of more than one demographic attribute. For example, in the example of Emma, she anonymized only her nationality; now suppose she did not register her nationality, age, and gender on Last.fm. She may think that it does not matter if her age is predicted but may think it is a big problem if her nationality and gender are predicted. In such a case, she tells Listener Anonymizer the two demographic attributes that she wants to preserve the anonymity of. The system shows an alert and recommends songs when at least one demographic does not satisfy not-first-anonymity. If more than one demographic attribute does not satisfy not-first-anonymity at the same time, the system needs to select songs that can recover not-first-anonymity for all of the demographic attributes by playing recommended the songs. When a user tells the system many demographics that she wants to preserve the anonymity of, alerts may frequently be displayed, and this makes it difficult for the user to enjoy listening to her favorite songs. Therefore, the user has to select demographic attributes that she really wants to preserve the anonymity of.

4. CAMOUFLAGING PLAY LOGS

In Section 3, we described various factors to be considered to realize Listener Anonymizer. In this section, based on these factors, we discuss the situation dealt with in this paper, give the problem definition, and propose a method for selecting songs for camouflaging play logs.

4.1 Problem Definition

In terms of the type of anonymity, we use not-first-anonymity because of its simplicity. If we can show the effectiveness of our proposed method in not-first-anonymity, we will deal with k-flat-anonymity in future work. As for the timing of selecting songs and camouflaging play logs, we camouflage the user's play log with as few songs as possible. That is, given user u's play log L_u that consists of m songs ($L_u = \{s_1, s_2, \cdots, s_m\}$ where s_i represents a song), we aim to anonymize u's confidential demographic attribute by selecting as few songs as possible. We assume our system knows the user's true demographic attributes. This assumption is reasonable because users will not hesitate to tell their demographics to the system if it is implemented as a stand-alone application as we described in Section 3.4. Finally, for preserving the anonymity of multiple demographics, since this paper deals with a new research problem, we consider single demographic anonymity as a first step. We are fully aware of the issue of multiple demographic anonymity; we leave this for future work.

Based on the above assumptions, our problem is defined as follows: "User u conceals an attribute value a_u in demographic d and wants to preserve not-first-anonymity regarding a_u. Given u's play log consisting of m songs, we verify if a_u satisfies not-first-anonymity. If it does not, we select as few songs as possible so that a_u can satisfy not-first-anonymity by playing them."

4.2 Demographic Prediction Method

To the best of our knowledge, the state-of-the-art method for prediction of users' demographic attributes of an online music service is the method proposed by Krismayer *et al.* [10]. They proposed a feature modeling approach. More specifically, given the users' play logs in the training data, they extract an artist name and the artist's tags as features for each song in each user's play log. Here, only the top 10,000 artists and top 10,000 tags in terms of the popularity in the training data are used to create feature vectors. They compute the weight for each feature in the form of TF-IDF values and create a feature vector for each user. The feature vectors, each of which has 20,000 dimensions, are reduced to 500 dimensions by principal component analysis (PCA) [6]. Finally, the classifier is built by using SVM. Since their evaluation results showed that the polynomial kernel achieved high prediction accuracy on average, we assume that using the SVM with the polynomial kernel is the state-of-the-art method. More details can be found in Krismayer *et al.* [10]. Once a classifier is built, given a user's play log, the classifier computes the probability distribution over demographic attribute values and outputs the attribute value that has the highest probability as the user's predicted attribute value. We implemented this prediction method by ourselves with reference to Krismayer *et al.* [10].

4.3 Song Selection Method

When Emma's nationality is predicted as French, Listener Anonymizer needs to select as few songs as possible that can anonymize her nationality. To achieve this, we aim to find songs that can largely increase the probability of the second-highest nationality (in this example, suppose Italian has the second highest probability). Since the feature vector corresponding to a song is compressed and the compressed vector is projected onto a new coordinate space by a polynomial kernel of SVM, it is difficult to find such songs based on the characteristics of an original 20,000-dimension vector. Instead, we assume such songs are played by users who are in the training dataset and are classified as Italian with high probabilities. From these users' play logs, we extract effective songs by using feature ablation analysis [1]. More formally, given L_u, we first compute the probability distribution over n attribute values by using the method in Section 4.2. If $p(a_u)$ is not the highest among them, we do not need to do anything. If $p(a_u)$ is the highest, we select songs as follows.

Suppose $a_j(\neq a_u)$ has the second-highest probability after a_u. Let $U = \{u_1^t, u_2^t, \cdots, u_q^t\}$ be a set of users in the training data. By developing the SVM classifier, user $u_i^t \in U$ has the probability $p(a_j, u_i^t)$ that represents the probability of u_i^t on a_j. From all users in U, we collect the top r users in terms of $p(a_j, u_i^t)$. Each user has her play log that consists of m songs. Suppose we remove the lth song from u_i^t's play log and compute the new probability of $p(a_j, u_i^t)$ by applying the SVM to the remaining $m-1$ songs (let the new probability be $p'(a_j, u_i^t)$). If the score of $p(a_j, u_i^t) - p'(a_j, u_i^t)$ is large, we assume that the lth song is essential to increase the probability of a_j. Based on this idea, we compute the score for each of the $r \times m$ songs and collect the top c corresponding artists based on the score. After collecting the top c artists, we randomly select one artist; then we randomly select one song of the artist's songs. By adding the song to L_u, we generate a camouflaged play log consisting of $m+1$ songs. We repeatedly select a song and add it to L_u until the camouflaged play log satisfies not-first-anonymity.

Table 1. Percentage of users who anonymize their demographic attributes. "✓" represents anonymization.

Age	Gender	Nationality	No. of users	%
✓	✓	✓	59,350	49.3
✓	✓		2,345	1.95
✓		✓	2,713	2.25
	✓	✓	454	0.377
✓			9,794	8.14
	✓		2,402	2.00
		✓	2,615	2.17
			4,0649	33.8

5. EXPERIMENTS

In this section, we carry out experiments to evaluate the effectiveness of our proposed method.

5.1 Dataset

We used the Last.fm dataset provided by Schedl [20]. As for the user's demographic attributes, this dataset includes age, gender, and nationality. Table 1 shows the numbers of users and the percentages for each of the combinations of confidential attributes, where "✓" indicates a confidential attribute. It can be observed that as many as 49.3% of users do not register any of their attributes, and 66.2% of them conceal at least one attribute. These statistics suggest the importance of preserving the user's demographic anonymity, though there might be other reasons. The dataset also includes users' play logs, each of which consists of the user ID, artist ID, track ID, and timestamp.

Following Krismayer *et al.* [10], we selected users who registered all of the three attributes, had equal to or more than 500 play logs, and had a nationality that was one of the 25 most common nationalities in terms of the number of users in the dataset. This gave us 32,991 users. We used 70% of them as training data and developed an SVM classifier. The remaining 30% of them were used as test data. Artists' tags were collected by using the Last.fm API [5]. The number of classes of each attribute is as follows. The nationality consists of 25 classes that correspond to the top 25 most common nationalities, the gender has two classes that are male and female, and following Schedl *et al.* [21], the age was divided into seven age groups ([6 - 17], [18 - 21], [22 - 25], [26 - 30], [31 - 40], [41 - 50], and [51 - 60]).

5.2 Methods Comparison

5.2.1 Settings

Our first research question is "Is our proposed method able to camouflage play logs with fewer songs than a base-

[5] `https://www.last.fm/api`

line method that randomly selects songs?" To answer this question, we count the number of songs selected by each method to preserve anonymity as follows. In this evaluation, for each user in the test dataset, we use the first 30 songs from the oldest songs in the play logs (*i.e.*, $m = 30$). For example, given the demographic attribute "nationality," we first compute each user's probability distribution over 25 nationalities when the 30 songs are played. We then sample 50 users whose nationality does not satisfy not-first-anonymity (*i.e.*, the user's nationality has the highest probability among 25 nationalities). Note that in this case, each user's play log in the training data also consists of 30 songs. Given a sampled user's play log consisting of 30 songs, we add a song selected by our proposed method and compute the new probability distribution for the 31 songs. If the probability of the user's nationality is not the highest among 25 nationalities, it means her nationality is anonymized and the song selection process ends; otherwise, we add a new song selected by our method and compute the probability distribution for the 32 songs. If the user's nationality is not anonymized even after selecting the additional 30 songs, we stop the song selection process. In this way, we count the number of selected songs for all of the 50 users. In this evaluation, the values of r and c were set to 3 and 1, respectively. Note that r is the number of users used for selecting candidate songs and c is the number of artists used for recommending songs as we described in Section 4.3. The random baseline method (hereafter, the random method) randomly selects a song from all the songs in the dataset and counts the number of songs in the same manner as described above.

In addition to the proposed method and the random method, we use a popularity-based baseline method (hereafter, the popularity method). Intuitively, in this method, if we want to decrease the probability of France, for example, we select a song that is not popular in France but is popular in the other 24 nationalities. To achieve this, we rank all artists in each nationality where an artist's score is the number of users who have listened to one of the artist's songs at least once. The artists are ranked in descending order of their score. When a user u's nationality a_u is not anonymized, the popularity method first selects an artist b^* where

$$b^* = \arg\max_{b \in B} \frac{\sum_{a_i \in A_d \setminus \{a_u\}} (\mathit{rank}(a_i, b) - \mathit{rank}(a_u, b))}{|A_d \setminus \{a_u\}|}.$$

In the equation, B is the set of all artists and $\mathit{rank}(a_i, b)$ represents the rank of artist b in nationality a_i. Finally, a song of b^* is selected and added to u's play log.

Note that in this evaluation, we used the same 50 users for all of the three methods for a fair comparison.

5.2.2 Results

The results are shown in Figure 1 where each bar represents the average number of selected songs over 50 users. It can be observed that our proposed method outperformed the other two methods in all attributes. In the "gender" attribute, even the proposed method selected as many as 19.68 songs on average. Since the "gender" attribute has only two classes (male and female), the probability tended to be strongly biased to one class. Thus, we presume that many songs were needed to fill the large gap. In the "nationality" attribute, the proposed method was especially effective: it selected less than one third of the songs selected by the random method.

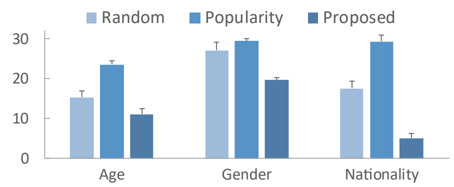

Figure 1. Comparison results between three methods. The y-axis is the average number of selected songs for camouflaging play logs. Error bars indicates the standard error.

The results of the popularity method were worse than those of the random method. This is because of the complexity of the demographic prediction method as described in Section 4.3. These results indicate that songs selected by the popularity method are rarely plotted to ideal points in the coordinate space created by an SVM polynomial kernel. Moreover, in the random method, a song that largely decreases the probability of the user's confidential attribute can be selected by chance. Because of these reasons, the random method outperformed the popularity method.

5.3 Parameter Effect

5.3.1 Settings

Remind that our method has a parameter c that determines how many artists we use from the result of the feature ablation, although we set c to 1 in Section 5.2. Our second research question is "What is the relation between the value of c in the proposed method and the number of selected songs?" To answer the question, we change the value of c from 1 to 10 and count the number of selected songs for each c. In each of the three demographic attributes, the same 50 sampled users were used for all of the c values.

5.3.2 Results

Figure 2 shows the results. In the "age" and "nationality" attributes, the number of selected songs decreases when c changes from 1 to 2 and the number is at a minimum when c is 2 or 3; then the number increases with the increase of c. In particular, in the "age" attribute, when c is 2, only 3.22 songs are required on average to camouflage play logs consisting of 30 songs. In the "gender" attribute, although the number of selected songs decreases when c changes from 1 to 2, the minimum score was 9.28 when c is 10. From these results, we can say that selecting songs only from the best artist in terms of feature ablation analysis does not lead to the best result.

In addition to the decrease of the number of selected songs for large c, the increase of c has another advantage. When c is 1, songs are always selected from one artist to anonymize an attribute value. This may enable a company that wants to predict users' demographic attributes to easily detect the camouflaged logs and predict the true attributes by removing the camouflaged logs. In contrast, when c is large, it becomes difficult to detect the

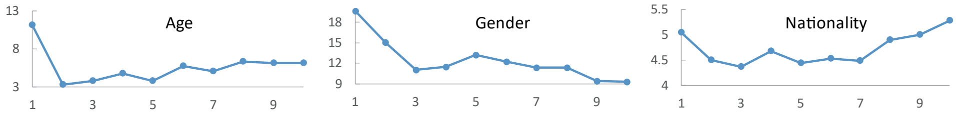

Figure 2. The relation between the value of c in the proposed method (x-axis) and the average number of selected songs for camouflaging play logs (y-axis).

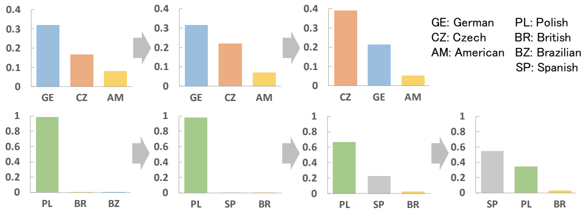

Figure 3. Examples of transition of probability distribution when two songs (top) and three songs (bottom) are selected for camouflaging play logs (y-axis: probability).

camouflaged parts in play logs. Moreover, by selecting songs from various artists, Listener Anonymizer may be able to expand the user's music taste while preserving her anonymity. Figure 2 shows that even when c is 10, our method can anonymize an attribute with much fewer songs than the random method in all attributes. It would be useful to enable a user to select the value of c while thinking about a trade-off between the number of songs and the diversity of selected songs.

Figure 3 shows examples of the transition of probability distribution when two or three songs are selected by our proposed method in the "nationality" attribute. The value of c was set to 3. For visibility, we show only the top three nationalities in terms of the probability. In the top example, the user's attribute is German. Listener Anonymizer can anonymize the user's attribute by selecting two songs. In the bottom example, although the initial probability distribution is strongly biased to the user's nationality (PL), this user can camouflage the play log by playing only three songs recommended by Listener Anonymizer.

6. DISCUSSION

In Section 5, we showed the effectiveness of the proposed method. However, since preserving demographic anonymity by camouflaging play logs is a quite new research theme, we discuss four important considerations.

6.1 Impact on Recommendation Accuracy

Since Listener Anonymizer camouflages play logs, it might degrade recommendation accuracy of music services. Although this paper dared to propose this controversial topic of research to give users an option of increasing the privacy and raise privacy issues in the MIR community, we are fully aware of the importance and usefulness of music recommendation to improve the user's music experience. We hope that our paper could contribute to discuss a diversity of options for music experiences while balancing privacy versus accuracy in music recommendation.

6.2 User's Taste in Music

In our method, the selected songs do not always match her taste in music. Even if those songs can camouflage the play logs, she might be reluctant to keep listening to the songs. Hence, it is beneficial to select songs by considering the user's taste in music. Many studies about song recommendation [11, 26, 28] and playlist generation [4, 13, 17] that can reflect the user's taste in music have been conducted. By introducing the methods proposed in these studies, we plan to propose a song selection method that can balance camouflaging the play logs and taste. That would also be beneficial to satisfy both anonymization and good recommendation.

Considering the user's taste has another advantage. If our method to camouflage play logs gains in popularity, companies that want to know the user's demographic attributes will try to predict them by removing songs that camouflage her play log. By selecting songs that match the user's taste, it becomes more difficult to detect songs that are played for camouflage.

6.3 Simulation of Multiple Prediction Methods

In our experiments, we assumed that the system knew that the method by Krismayer *et al.* [10] is used to predict the user's demographic attributes. However, we cannot always know the prediction method in advance. When we do not know it, one strategy is to prepare multiple possible prediction methods and simulate them one at a time. An alert is issued when more than v methods detect that the user's demographic attribute can be predicted with high accuracy. For small v, the degree of anonymity preservation is high but alerts are often issued and vice versa for large v. It would be useful for a user to be able to set the value of v according to the degree of anonymity she requires.

6.4 Real-time Monitoring of a Play Log

In our experiments, the number of songs in a given play log was set to 30. Hence, all logs in training data also consisted of 30 songs. However, this assumption is not sufficient to monitor the user's played songs and recommend songs at the right time as we described in Section 3.3. This is because, when a user plays her first song, there is no play log in the training data consisting of only one song and we cannot correctly compute the probability distribution for the song. To solve this problem, we need to develop classifiers for various values of l, where l is the number of songs included in a play log.

7. CONCLUSION

In this paper, we proposed Listener Anonymizer that can preserve the user's demographic anonymity by camouflaging her play log. Our experimental results show the effectiveness of our proposed method to select as few songs as possible. For example, in the "age" attribute, 15.3 songs were selected by the random method, while only 3.22 songs were selected by our method. Since this paper proposed a new concept, there are many remaining issues to be addressed as we discussed in Section 3 and 6. We plan to tackle them one by one and make Listener Anonymizer more flexible and useful.

8. ACKNOWLEDGMENTS

This work was supported in part by JST ACCEL Grant Number JPMJAC1602, Japan.

9. REFERENCES

[1] Eugene Agichtein, Ryen W. White, Susan T. Dumais, and Paul N. Bennet. Search, interrupted: Understanding and predicting search task continuation. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'12, pages 315–324, 2012.

[2] John D. Burger and John C. Henderson. An exploration of observable features related to blogger age. In *In Computational Approaches to Analyzing Weblogs: Papers from the 2006 AAAI Spring Symposium*, pages 15–20, 2006.

[3] William M. Campbell, Douglas E. Sturim, and Douglas A. Reynolds. Support vector machines using gmm supervectors for speaker verification. *IEEE Signal Processing Letters*, 13(5):308–311, 2006.

[4] Arthur Flexer, Dominik Schnitzer, Martin Gasser, and Gerhard Widmer. Playlist generation using start and end songs. In *Proceedings of the 9th International Conference on Music Information Retrieval*, ISMIR'08, pages 173–178, 2008.

[5] Sumit Goswami, Sudeshna Sarkar, and Mayur Rustagi. Stylometric analysis of bloggers' age and gender. In *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media*, ICWSM'09.

[6] Harold Hotelling. Analysis of a complex of statistical variables with principal components. *Journal of Educational Psychology*, 24:417–441 and 498–520, 1933.

[7] David Huffaker and Sandra Calvert. Gender, identity, and language use in teenage blogs. *Journal of Computer-Mediated Communication*, 10(2), 2005.

[8] Rosie Jones, Ravi Kumar, Bo Pang, and Andrew Tomkins. "I know what you did last summer": Query logs and user privacy. In *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, CIKM'07, pages 909–914, 2007.

[9] Margaret L. Kern, Johannes C. Eichstaedt, H. Andrew Schwartz, Gregory Park, Lyle H. Ungar andDavid J. Stillwell, Michal Kosinski, Lukasz Dziurzynski, and Martin E. P. Seligman. From "sooo excited!!!" to "so-proud": Using language to study development. *Developmental psychology*, 50(1):178–188, 2014.

[10] Thomas Krismayer, Markus Schedl, Peter Knees, and Rick Rabiser. Prediction of user demographics from music listening habits. In *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*, CBMI'17, pages 8:1–8:7, 2017.

[11] Dawen Liang, Minshu Zhan, and Daniel P. W. Ellis. Content-aware collaborative music recommendation using pre-trained neural networks. In *Proceedings of the 16th International Society for Music Information Retrieval Conference*, ISMIR'15, pages 295–301, 2015.

[12] Jen-Yu Liu and Yi-Hsuan Yang. Inferring personal traits from music listening history. In *Proceedings of the 2nd International ACM Workshop on Music Information Retrieval with User-centered and Multimodal Strategies*, MIRUM'12, pages 31–36, 2012.

[13] Beth Logan. Content-based playlist generation: Exploratory experiments. In *Proceedings of the 3rd International Conference on Music Information Retrieval*, ISMIR'02, pages 295–296, 2002.

[14] Huina Mao, Xin Shuai, and Apu Kapadia. Loose tweets: An analysis of privacy leaks on twitter. In *Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society*, WPES'11, pages 1–12, 2011.

[15] Dong-Phuong Nguyen, Rilana Gravel, Rudolf Berend Trieschnigg, and Theo Meder. "How old do you think I am?": A study of language and age in twitter. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*, ICWSM'13, pages 439–448, 2013.

[16] Dong-Phuong Nguyen, Rudolf Berend Trieschnigg, A. Seza Dogruoz, Rilana Gravel, Mariet Theune, Theo Meder, and Franciska M.G. de Jong. Why gender and age prediction from tweets is hard: Lessons from a crowdsourcing experiment. In *Proceedings of the 25th International Conference on Computational Linguistics*, COLING'14, pages 1950–1961, 2014.

[17] Elias Pampalk, Tim Pohle, and Gerhard Widmer. Dynamic playlist generation based on skipping behavior. In *Proceedings of the 6th International Conference on Music Information Retrieval*, ISMIR'05, pages 634–637, 2005.

[18] Maarten Sap, Gregory Park, Johannes C. Eichstaedt, Margaret L. Kern, David Stillwell, Michal Kosinski, Lyle H. Ungar, and H. Andrew Schwartz. Developing age and gender predictive lexica over social media. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, EMNLP'14, pages 1146–1151, 2014.

[19] Pierre Saurel, Francis Rousseaux, and Marc Danger. On the changing regulations of privacy and personal information in mir. In *Proceedings of the 15th International Society for Music Information Retrieval Conference*, ISMIR'14, pages 597–602, 2014.

[20] Markus Schedl. The lfm-1b dataset for music retrieval and recommendation. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, ICMR'16, pages 103–110, 2016.

[21] Markus Schedl, David Hauger, Katayoun Farrahi, and Marko Tkalčič. On the influence of user characteristics on music recommendation algorithms. In *Proceedings of the 37th European Conference on Information Retrieval*, ECIR'15, pages 339–345, 2015.

[22] Alexandra L. Uitdenbogerd and Ron G. van Schyndel. A review of factors affecting music recommender success. In *Proceedings of the 3rd International Society for Music Information Retrieval Conference*, ISMIR'02, pages 204–208, 2002.

[23] Gabriel Vigliensoni and Ichiro Fujinaga. Automatic music recommendation systems: Do demographic, profiling, and contextual features improve their performance? In *Proceedings of the 17th International Society for Music Information Retrieval Conference*, ISMIR'16, pages 94–100, 2016.

[24] Yuan Wang, Yang Xiao, Chao Ma, and Zhen Xiao. Improving users' demographic prediction via the videos they talk about. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, EMNLP'16, pages 1359–1368, 2016.

[25] Ming-Ju Wu, Jyh-Shing Roger Jang, and Chun-Hung Lu. Gender identification and age estimation of users based on music metadata. In *Proceedings of the 15th International Society for Music Information Retrieval Conference*, ISMIR'14, pages 555–560, 2014.

[26] Zhe Xing, Xinxi Wang, and Ye Wang. Enhancing collaborative filtering music recommendation by balancing exploration and exploitation. In *Proceedings of the 15th International Society for Music Information Retrieval Conference*, ISMIR'14, pages 445–450, 2014.

[27] Billy Yapriady and Alexandra L. Uitdenbogerd. Combining demographic data with collaborative filtering for automatic music recommendation. In *Knowledge-Based Intelligent Information and Engineering Systems*, volume 3684 of *LNCS*, pages 201–207. 2005.

[28] Kazuyoshi Yoshii, Masataka Goto, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno. Hybrid collaborative and content-based music recommendation using probabilistic model with latent user preferences. In *Proceedings of the 7th International Conference on Music Information Retrieval*, ISMIR'06, pages 296–301, 2006.