Experience Individualization on Online TV Platforms through Persona-based Account Decomposition

Payal Bajaj
Adobe Research Labs
Bengaluru, India
pabajaj@adobe.com

Sumit Shekhar
Adobe Research Labs
Bengaluru, India
sushekha@adobe.com

ABSTRACT

Online TV has seen rapid growth in recent years, with most of the large media companies broadcasting their linear content online. Access to the online TV accounts is protected by an authentication, and like the traditional cable TV subscription, users in the same household share the online TV credentials. However, as the standard data collection techniques have capability to collect only account level information, online TV measurements fail to capture individual level viewing characteristics in shared accounts. Thus, individual profile identification and experience individualization are challenging and difficult for online TV platforms. In this paper, we propose a novel approach to decompose online TV account into distinct personas sharing the account through analyzing viewing characteristics. A recommendation algorithm is then proposed to individualize the experience for each persona. Finally, we demonstrate the usefulness of the proposed approach through experiments on a large online TV database.

Keywords

Online TV measurement; Experience Individualization; Cluster Analysis; Recommendation; Personalization

1. INTRODUCTION

The online TV industry has seen a massive growth in the last few years. The users can now access TV program content over the Internet using a multitude of devices at any time unlike the traditional TV, where users watch as per channel schedules. Online TV is also different from the traditional television as it is built to provide a personalized TV experience to its users.

The access to the online TV platforms usually requires user authentication, and like the traditional TV, it is meant to be consumed by the members of the family. In this scenario, the multitude of analysis methods available on online platforms, lose the ability to track individual user behavior. This limits the ability of to create an individualized online TV experience for its viewers. For instance, consider a situation where the account is being shared between multiple members in the household with varied channel preferences. The father in the house might prefer *business* and *news* channels, the mother might prefer *travel* and *cooking* channels, the elder children may prefer *sports* channels whereas the younger kids might prefer *cartoon* and *anime* channels. Hence, decomposition of online TV accounts into individual viewing patterns is important to achieve the personalized TV experience.

This is a challenging problem due to ubiquitous access to online TV content available to users. Viewership was restricted to linear shows and followed a specific routine in traditional TV, but in online TV, users don't exhibit routine and periodic behavior as the content availability is not restricted like before. Therefore, it is difficult to disambiguate individual user viewing characteristics using routine or periodicity from online TV viewing logs. In this paper, we demonstrate that routine viewing patterns don't exist and propose an approach to individualize experience of users using shared accounts in online TV that does not rely on habitual and routine pattern of the programs.

To this end, we assume that users sharing an account have distinct video content and genre-driven preferences for program channels. We propose a novel method for persona identification to disambiguate between such users. Personas are defined as common viewing patterns observed in online TV, which we infer through mining user access logs. We further describe a recommendation algorithm to exploit the persona decomposition of online TV accounts and demonstrate its superiority over routine, like time of the day, decomposition on a large online TV dataset.

The remainder of the paper is organized as follows. Section 2 describes the prior literature related to user profile individualization in online TV domain. The problem statement and the proposed approach are discussed in Section 3. The dataset and experiments are then summarized, followed by results and evaluation in Section 4. The paper is concluded in Section 5.

2. RELATED WORK

In this section, we discuss the prior literature for account decomposition and recommendation on online TV platforms.

2.1 Account Decomposition

Prior works in account disambiguation have focused on extracting routine behavior to decompose shared accounts. Iguchi *et al* [5] proposed a technique for account decompo-

MM '16, October 15–19, 2016, Amsterdam, The Netherlands.

DOI: http://dx.doi.org/10.1145/2964284.2967221

sition based on collective hourly viewing patterns. Wang et al [6] considered that user consuming behavior is periodic in IP-TV services and proposed a user identification algorithm based on captured preferences in continuous periods of non-overlapping sub-periods. User identification was also studied by Zhang et al. [4], where accounts were decomposed based on a subspace-based clustering method.

2.2 User Profiling using TV logs

User profiling plays a very important role for recommendation using online TV logs [8]. Online TV logs describe the content consumption behavior by the consumers, which can be used to infer their interests on watched TV programs.

Aroyo *et al* [1] presented a distributed personalization based method for TV personalization. A comprehensive context-sensitive user modeling method through monitoring user's actions was proposed by Pieter *et al* [2]. However, the context-sensitive user model created thus is independent of platform the content is being viewed on. For instance, the system assumes when a user sitting at his computer adds a program to his favorites, this will have a direct effect on his recommendation on the set top box application. But for online TV, users sharing the account would have different content preferences, hence this might not work well.

Zhiwen *et al* [3] proposed a method to construct a common profile for multiple viewers based on total distance minimization. They prove that merging can reflect the preferences of the majority of members within the group and recommendation strategy is effective for multiple viewers watching TV together. In contrast, we are trying to decompose the users within a shared account. Further, [3] requires multiple viewers to simultaneously log-in in order to identify who are watching the TV programs together.

Contributions: The contributions of this paper are -

- We propose a novel method to identify distinct personas sharing an account based on detecting the common viewing patterns across online TV users. We further demonstrate that habitual and routing watching patterns, assumed in prior literature, may not extend to online TV.
- We present a persona-based experience individualization method and demonstrate that it works better than context-based profiling method on a large online TV access logs dataset.

3. PERSONA-BASED INDIVIDUALIZATION

In this section, we describe the proposed approach for decomposing online TV accounts into personas followed by persona-based experience individualization. We also describe user profile and context-based profiling methods, against which we evaluate the proposed algorithm.

3.1 Notations

We denote the set of access logs by L, the set of program channels by Ch and the set of accounts in the dataset by A. Let the the total number of accounts be n, i.e., $n = |A|$. For each account $a \in A$, our problem is to identify distinct personas, a_p, sharing the account.

3.2 Persona Identification

We propose a method of account decomposition through persona identification. It distinguishes various users sharing

Algorithm 1 Proposed approach for persona identification from online TV access logs.

Input: L, A, Ch
Output: Distinct Personas $\{a_p\}\ \forall\ a \in A$
a. Create Frequency Matrix F
1: **for each** log $l \in L$ **do**
2: $a = Account(l)$
3: $ch = Channel(l)$
4: $F[a, ch] = F[a, ch] + 1$
5: **end for**
b. Create Correlation Matrix C
6: **for each** $ch_1 \in Ch$ **do**
7: $ch_1 = F[, ch_1]$
8: **for each** $ch_2 \in Ch$ **do**
9: $ch_2 = F[, ch_2]$
10: $corr(ch_1, ch_2) =$
11: $C[ch_1, ch_2] = CosineSimilarity(ch_1, ch_2)$
12: **end for**
13: **end for**
c. Channel Clustering
14: $Ch_{Clust} = HierarchicalClustering(C)$
15: $Clust_{Indx} = Indexes(Ch_{Clust})$
d. Splitting Accounts into Clusters
16: **for each** $a \in A$ **do**
17: Initialize set of cluster indexes, $a_c = \{\}$
18: **for each** $ch \in Ch$ **do**
19: **if** $F[a, ch] > 0$ **and** $Clust_{Indx} \notin a_c$ **then**
20: $a_c = a_c \cup Clust_{Indx}(ch)$
21: **end if**
22: **end for**
23: **end for**
e. Get Personas using Frequent Cluster Mining
24: $Clust_{Dataset} = \{\cup a_c \forall a \in A\}$
25: $Persona = Apriori(Clust_{Dataset}, support_{threshold})$
f. Decomposing Account into Personas
26: **for each** $a \in A$ **do**
27: $a_p = SetCover(a_c, Persona)$
28: **end for**

an account by creating personas with distinct preferences. We define personas as the top common viewing patterns among the online TV accounts. Further, we assume that the viewing characteristics of individual users are governed by their genre preferences. For example, kids in the household might prefer to watch cartoons, teenagers might prefer movies and TV series whereas an elderly person might spend more time on news. We arrive at these characteristic viewing patterns through mining access logs of millions of online TV users. The steps for arriving at the personas are described below. Algorithm 1 summarizes the proposed method.

3.2.1 *Similarity-based Channel Clustering*

The algorithm begins with identifying the dominant viewing patterns over all accounts. We analyze all the channels and compute pairwise cosine similarity between them. A frequency matrix F is created which counts the videos viewed by each account for every channel (Lines $1-5$ in algorithm 1). A pairwise similarity matrix C is created (Lines $6-13$) using F (Line 10). Then we group them into clusters based on these similarity values. Clustering is performed on this matrix (Lines $14-15$) using hierarchical clustering

[12]. These clusters reflect the dominant channel viewing characteristics across all accounts.

3.2.2 Account Splitting

We split accounts into the clusters obtained in the previous step. Each account a is decomposed on the basis on which cluster's channels were viewed by the account (Lines $16-23$). If a channel ch is viewed by account a, the corresponding cluster that ch belongs to is added to patterns for accounts a, i.e., a_c.

3.2.3 Frequent Cluster Mining to obtain Personas

Next we identify top viewing patterns through running frequent cluster mining on the channel clusters, a_c identified for each account. We create a dataset, $Clust_{Dataset}$ of these clusters for all accounts (Line 24), and run Apriori algorithm [7] with a specified support threshold, $support_{threshold}$ to find top viewed cluster combinations (Line 25). These cluster combinations are then identified as unique personas, $Persona$. Note that these personas represent the common TV viewing patterns as envisioned in Section 1. Further, each persona may include multiple clusters, like news and sports, which may not have been fallen in the same cluster due to the presence of multiple popular sports as well as news channels online.

3.2.4 Decomposing Account into Personas

We replace the clusters a_c identified for each account a by the personas, $Persona$, using the greedy set cover approximation algorithm [10] (Lines $26-28$). For example, if an account is decomposed into three clusters {sports, cartoons, news, fashion}, and personas are identified as the clusters {sports, cartoons}, {fashion} and {cartoons, news}, then this account has three personas using the above mentioned greedy approach.

3.3 Experience Individualization

We assume that accounts are shared by multiple personas, but a TV session is usually driven by one persona. In other words, we assume that even when multiple users with varied interests may share the account, a session involves viewing channels catering to a particular persona. For this paper, we have defined session as a period of channel viewing activities until no activity is observed for an hour. For instance, the kids in the house would use the shared household account to watch cartoon channels in a session, whereas the sports enthusiasts would use the same account to view a match at a continuous stretch. Therefore, our hypothesis is that a session is usually driven by persona-based preferences. We validate this through a persona-based recommendation system.

To implement persona-based recommendation, we first recognize the persona to be leveraged for recommendation. We use the first channel viewed by the account in a session to identify the persona that owns the session. We have created persona in a way that every persona maps to a cluster of channels. These channels are referred to as the profile for that persona. We use this persona profile to recommend additional channels to the account through the session.

3.4 Comparisons

To evaluate persona-based recommendations, we create two account-based profiles. First, we create a profile to capture the overall account behavior. Second, we create contextual profiles of the accounts to study the impact of context, specifically, time of watching on recommendation. The account and contextual profiles are created by a normalized count of channel views from the training data. For a TV user account a, the account profile $AccProfile_a$ is computed in equation 1 below. F represents the frequency matrix as before.

$$AccProfile_a = \frac{\{F[a, ch_1], F[a, ch_2], \cdots, F[a, ch_n]\}}{\sum_{ch \in Ch} F[a, ch]} \quad (1)$$

To create contextual profiles, we divide the channel views from the accounts to create two time-dependent profiles of the account - day and night profiles. The two time-dependent profiles of the account are calculated using Equation 1, after separating the channels viewed at day and night. For recommendations using account-based profile, we use kNN algorithm [11] to identify the k most similar accounts based on that account's overall profile. For contextual profile, time of the session is used to identify the same separately for day and night profiles. We have used traditional cosine similarity for this purpose. The channels viewed by these top-k online TV accounts are then recommended.

4. EXPERIMENTS AND RESULTS

In this section, we present results on account decomposition based on persona identification and evaluate the performance of the persona based recommendation system.

4.1 Dataset

The experiments are performed on the access logs of during the period between December 1, 2014 till December 31, 2014. The logs provide information about the channels viewed, the time and day of viewing and devices used by online TV accounts. The dataset contains about 5 million accounts, with access to over 110 channels. The set of channels spans a variety of genres including sports, animation, fashion, movies, and news, etc.

4.2 Viewing Patterns on online TV

In this section, we analyze the online video viewing behavior of all the accounts. We demonstrate that the viewing behavior for online TV content doesn't follow routinely periodic patterns for majority of the accounts.

TV shows usually follow a weekly pattern, therefore for channels under consideration, we measure the weekly engagement of the accounts. We define the most engaging day for an online TV account to be the day of the week on which it was the most active, and then count the number of weeks the particular account was engaged the most on that day of the week. Previous works in this area [5] [6] have assumed weekly engagement as an important criteria for account disambiguation. However, as the Figure 1 shows, less than 13% of the accounts exhibit routine engagement for more than 2 weeks across different days of the week. Thus it is difficult to disambiguate accounts based on routine behavior. Note that this analysis is done for all the channels without hourly granularity. Adding these constraints will only increase the sparsity of the patterns. Therefore, we can not use weekly habitual viewing characteristics to decompose TV accounts because such patterns are not exhibited by the data.

Therefore, it is difficult to decompose online TV accounts based on habitual patterns as opposed to traditional TV.

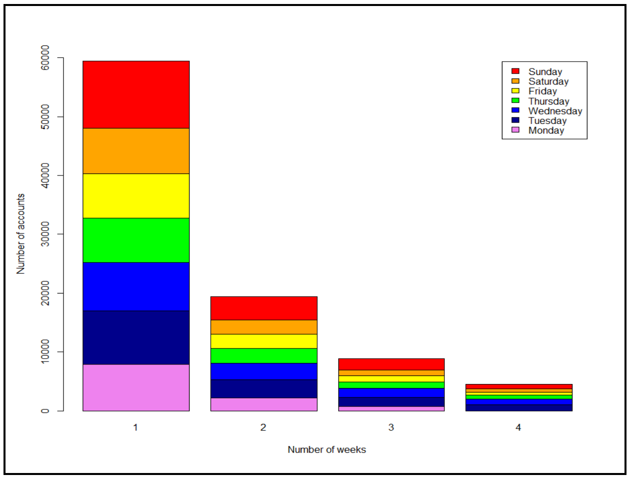

Figure 1: Plot of number of TV accounts with maximum engagement on a day of the week versus the number of active weeks for the account.

4.3 Experimental Set-up

In this section, we evaluate the proposed persona based recommendation algorithm. We took the logs for the first 20 days as training, and then evaluated the model for the remaining days of the month. For clustering, we have chosen the height of hierarchical clustering algorithm to be 1.4 empirically. Further, we keep the $support_{threshold}$ at 0.01 to discover representative personas for online TV accounts. Also, the value of k to create account and contextual profiles is chosen empirically to be 5.

We compare the channels recommended using account profiles, contextual profiles and persona profiles. For contextual profiles, we have used one of the day or night profiles depending on the time of the day in the test dataset. We report the recall and accuracy metrics [1] for comparing these algorithms.

4.4 Results

Figure 2 shows the output clusters of TV channels watched together (Line 14 in Algorithm 1). It can be observed that channels within many clusters have similar genre content. The final personas (Line 25) also reflect similar output (final persona outputs are not illustrated due to paucity of space). Thus, our assumption that distinct personas within an account are driven by content holds true.

The results of the recommendation algorithm are presented in Table 1. We took profiles with multiple discovered personas for testing the recommendations. The results are reported across accounts with 2, 3 and 4 discovered personas, along with the size of respective sets. It can be seen that the recall of persona-based recommendation outperforms the

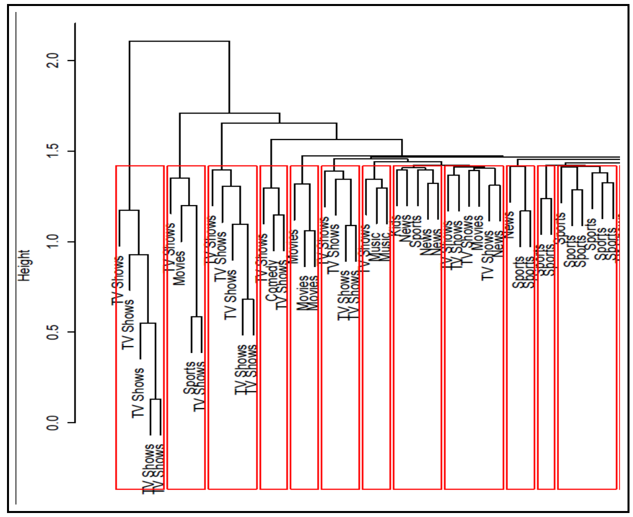

Figure 2: Clusters of different channels, showing their respective content genre, discovered through the proposed algorithm.

Algorithm	No. of Personas (Total count)	Recall	Accuracy
Account-based	2 (381,000)	0.4057	0.9872
Context-based	2 (381,000)	0.3945	0.9867
Persona-based	2 (381,000)	0.6973	0.9654
Account-based	3 (54,000)	0.3578	0.9852
Context-based	3 (54,000)	0.3422	0.9844
Persona-based	3 (54,000)	0.6226	0.9621
Account-based	4 (10,170)	0.2952	0.9826
Context-based	4 (10,170)	0.2948	0.9822
Persona-based	4 (10,170)	0.5878	0.9601

Table 1: Results for different channel recommendation algorithms.

account and contextual profile-based methods, with only a slight decrease in accuracy. Contextual profiles do not improve the results at all. This validates our claim that online TV behavior is significantly different from traditional TV watching patterns.

5. CONCLUSION AND FUTURE WORK

In this paper, we present the problem of account disambiguation for online TV accounts. The patterns of online TV account usage make it difficult for time and routine based account decomposition. We present a novel approach which decomposes the TV accounts into distinct personas sharing the account through analyzing the channel viewing patterns. We further built a persona-based channel recommendation system, and demonstrate that it performs better than user-profile and contextual recommendation system on a large dataset of online TV logs.

Exploration of different methods to evaluate account decomposition into personas is a future work, as obtaining the ground truth is hard. Further, testing persona-based decomposition with advanced algorithms would highlight the significance of the proposed framework.

[1]

$$Recall = \frac{True\ Positive(TP)}{True\ Positive(TP)\ +\ False\ Negative(FN)}$$

$$Accuracy = \frac{True\ Positive(TP)\ +\ True\ Negative(TN)}{Totalnumber\ of\ channels}$$

6. REFERENCES

[1] Aroyo, L. M. ”Distributed personalization: Bridging digital islands in museum and interactive TV.” ERCIM News 72 (2008): 40.

[2] Bellekens, Pieter, Geert-Jan Houben, Lora Aroyo, Krijn Schaap, and Annelies Kaptein. ”User model elicitation and enrichment for context-sensitive personalization in a multiplatform TV environment.” Proceedings of the 7th European Conference on European interactive Television Conference. ACM, 2009.

[3] Yu, Zhiwen, Xingshe Zhou, Yanbin Hao, and Jianhua Gu. ”TV program recommendation for multiple viewers based on user profile merging.” User modeling and user-adapted interaction 16.1 (2006): 63-82.

[4] Zhang, Amy, Nadia Fawaz, Stratis Ioannidis, and Andrea Montanari. ”Guess who rated this movie: Identifying users through subspace clustering.” arXiv preprint arXiv:1208.1544 (2012).

[5] Iguchi, Koichi, Yoshinori Hijikata, and Shogo Nishida. ”Individualizing user profile from viewing logs of several people for TV program recommendation.” Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication. ACM, 2015.

[6] Wang, Zhijin, Yan Yang, Liang He, and Junzhong Gu. ”User Identification within a Shared Account: Improving IP-TV Recommender Performance.” Advances in Databases and Information Systems. Springer International Publishing, 2014.

[7] Agrawal, Rakesh, and Ramakrishnan Srikant. ”Fast algorithms for mining association rules.” Proc. 20th International Conference on Very Large Databases, VLDB. Vol. 1215. 1994.

[8] Vèras, Douglas, et al. ”A literature review of recommender systems in the television domain.” Expert Systems with Applications 42.22 (2015): 9046-9076.

[9] Han, Jiawei, Wan Gong, and Yiwen Yin. ”Mining Segment-Wise Periodic Patterns in Time-Related Databases.” KDD. 1998.

[10] Chvatal, V. A. ”Greedy Heuristic for the Set-Covering Problem.” Mathematics of Operations Research Vol. 4, No. 3 (Aug., 1979), pp. 233-235

[11] Cover, Thomas M., and Peter E. Hart. ”Nearest neighbor pattern classification.” IEEE Transactions on Information Theory, 13.1 (1967): 21-27.

[12] Johnson, Stephen C. ”Hierarchical clustering schemes.” Psychometrika 32.3 (1967): 241-254.