Time Feature Selection for Identifying Active Household Members

Pedro G. Campos[a,b], Alejandro Bellogín[a], Fernando Díez[a], Iván Cantador[a]
pgcampos@ubiobio.cl, {alejandro.bellogin, fernando.diez, ivan.cantador}@uam.es

[a]Universidad Autónoma de Madrid
Escuela Politécnica Superior
28049 Madrid, Spain

[b]Universidad del Bío-Bío
Departamento Sistemas de Informacion
4081112 Concepción, Chile

ABSTRACT

Popular online rental services such as Netflix and MoviePilot often manage household accounts. A household account is usually shared by various users who live in the same house, but in general does not provide a mechanism by which current active users are identified, and thus leads to considerable difficulties for making effective personalized recommendations. The identification of the active household members, defined as the discrimination of the users from a given household who are interacting with a system (e.g. an on-demand video service), is thus an interesting challenge for the recommender systems research community. In this paper, we formulate the above task as a classification problem, and address it by means of global and local feature selection methods and classifiers that only exploit time features from past item consumption records. The results obtained from a series of experiments on a real dataset show that some of the proposed methods are able to select relevant time features, which allow simple classifiers to accurately identify active members of household accounts.

Categories and Subject Descriptors

I.5.1 [**Pattern recognition**]: Models

General Terms

Algorithms, Performance, Experimentation

Keywords

Household Member Identification, Time Features, Feature Selection, Recommender Systems

1. INTRODUCTION

Popular online rental services such as Netflix [1] and MoviePilot [2] often manage household accounts, that is, accounts shared by several users who usually live in the same house. For a particular household account, its members do not always access the service and consume together offered items at the same time. Moreover, in general, there is no mechanism to identify current active users, i.e., users accessing the service at a particular time. These issues lead to considerable difficulties and limitations for providing an effective personalized assistance, e.g. by making personalized item recommendations [3, 6]. The identification of active household members, defined as the discrimination of which users from a given household are interacting with a system (e.g. an on-demand video service), is thus an interesting challenge for the recommender systems (RS) research community.

To address such challenge, we may follow two main types of strategies. A first type of strategies would attempt to increase the diversity of recommendations [8], aiming to cover the heterogeneous range of preferences of the different members in a household. A second type of strategies, on the other hand, would attempt to identify the active household members for which recommendations have to be delivered. We advocate for and focus on the second type of strategies since it lets making more accurate item recommendations, by only using preferences of active members and discarding preferences of other, non present members [3].

In this paper, we formulate the identification of the active household members as a classification problem, which we address by exploiting only time information associated to the users' interactions (e.g. ratings) in the system. Extending previous work [1], we explore and evaluate different representations of time information, along with several global and local time feature selection methods. By doing so, we aim to address the following research questions: **RQ1**, does time information alone provide enough discriminating power for the active household member identification task? And **RQ2**, are there meaningful differences between using global and local methods for time feature selection? The results obtained from experiments on a real dataset show that some of the proposed methods are able to select informative time features, which are exploited by simple classifiers to accurately identify active members of household accounts.

2. RELATED WORK

In the Recommender Systems research field, there have been recent efforts to address the identification of individual users from household accounts [3, 6]. Specifically, in the 2011 edition of the Context-Aware Movie Recommendation Challenge [6], CAMRa, participants had to identify which members of particular households were responsible for a number of events –interactions with the system in the form of ratings. The contest provided a training dataset with meta-information about ratings in a movie recommender system, such as the household members who actually provided the ratings, and the associated timestamps. The contest's goal was to identify the users who had been responsible for cer-

[1] www.netflix.com
[2] www.moviepilot.com

CIKM 2012, October 29–November 2, 2012, Maui, HI, USA.

tain events (ratings) whose household and timestamp were given in a test dataset. This task was assumed to be equivalent to the task of identifying active users requesting recommendations at a particular time.

The winners of the 2011 CAMRa challenge [2] so as some other participants (e.g. [1, 7]) exploited several time features derived from the available event timestamps. The used features showed different temporal rating habits of users in a household, regarding the day of the week, the hour of the day, and the month of the year, when the users had rated items. Other time features that can be derived from timestamps, such as the period of the day (e.g. morning, noon, afternoon, evening) and the period of the week (workday, weekend) were not analyzed in those works. Some of the proposed models also exploited other non time related features, such as the specific rating values. In this paper, we focus on methods that rely exclusively on time features derived from timestamps. This prevents our approach to be restricted to specific domain or application dependent information (e.g. explicit vs. implicit user preferences), and let us to address a generic recommendation setting.

Other works have also dealt with problems raised by the use of shared user accounts. In the context of the Netflix Prize competition, Koren [4] discusses some difficulties for RS that could emerge from the use of household accounts (note that the well-known Netflix dataset in fact contains household identifiers, not user identifiers). In that work, Koren proposes a temporal recommendation model that assumes the existence of a drifting *meta-user* associated with each household account. Similarly, Kabutoya et al. [3] aim to identify *latent users* sharing an account, by using a probabilistic topic model. The above works use household-level training data, i.e., data where it is unknown which users compose a household and which household members really provided particular (training) ratings; and aim to improve recommendation accuracy over withheld test ratings. With this respect, these works differs from the one presented here, in the sense that they focus on detecting latent preference patterns within account user profiles, in order to improve the final performance of a certain recommender system. We propose, on the contrary, to model knowledge about such patterns independently from any recommender. Thus, in our approach, once the active members of household accounts are identified, any recommendation algorithm could be used. In this way, we believe that recommended items would better fit the active users' preferences.

3. IDENTIFYING ACTIVE HOUSEHOLD MEMBERS BASED ON TIME FEATURES

In the 2011 CAMRa challenge, the approach that achieved the highest performance on the household member identification task was based on the classification of user patterns, described by feature vectors that include time context information [1, 2]. We can formalize that approach as follows. Let us consider a set of events $E = \{e_1, e_2, ..., e_m\}$ and a set of users $U_{\mathbb{h}} = \{ u_{1,\mathbb{h}}, u_{2,\mathbb{h}}, ..., u_{n,\mathbb{h}}\}$ within a household $\mathbb{h}$, such that each event e_i is associated to one, and only one, user $u_{j,\mathbb{h}}$. Also, let us consider that each of these events is described by means of a feature vector, called X_{e_i}. The question to address is whether it is possible to determine which user is associated to an event e_i once (some) components x_{e_i} of its feature vector X_{e_i} are already known. In this paper, events correspond to instances of user ratings, and feature vectors correspond to time context representations of the events.

Table 1 shows the time features explored in this paper. Aiming to estimate the discrimination power of these features, we used the well-known *Kullback-Leibler Divergence* (KLD) [5], which lets us to measure the divergence between pairs of users in a household, regarding the probability distribution of the features for the users in each household. Higher values correspond to more divergent probability distributions, and can be interpreted as having users in households with differentiated habits with respect to the corresponding time features. In the table the features are sorted in descending order by the average KLD value computed over all pairs of users in each household. Hence, the best discriminant features are the **Date** (D), the **Day of Week** (W), and the **Hour of day** (H).

Table 1: Proposed time features

Time feature	Domain	KLD
D: Date	$1, 2, \cdots$,#days in training set	5.79
W: Day of Week	$1, 2, \cdots, 7$	4.56
H: Hour of day	$0, 1, \cdots, 23$	4.53
P_d: Period of day	$morning, noon, afternoon, evening$	2.28
P_w: Period of week	$workday, weekend$	1.79
M: Meridian	AM, PM	1.47
M_h: Minute of hour	$0, 2, \cdots, 59$	0.97
Q_h: Quarter of hour	$1, 2, 3, 4$	0.70
M_y: Month of year	$1, 2, \cdots, 12$	0.36

The use of KLD as a predictor of the discrimination power of a time feature in the household member identification task requires to be confirmed experimentally. Additionally, the use of feature vectors including different combination of time features may have diverse impact on the identification performance. In order to test the true discrimination power of the explored time features, we use distinct classification methods to identify the user associated to an event in a given household. In Section 3.1 we present global methods that use a fixed set of features for all the households, and in Section 3.2 we present local methods that dynamically select the most valuable features for each household.

3.1 Global Classification Methods

In this section we present several methods that use a fixed set of time features in the classification task, i.e., they use the same set of features over all the households. The first evaluated method is the *A priori* model proposed in [1]. This method computes probability distribution functions, which represent the probabilities that users are associated to particular events, and uses computed probabilities to assign a score to each user in a household given a new event. More specifically, we compute the probability mass function (PMF) of each feature given a particular user, restricted to the information related with that user's household, that is, $\{p(X = x_i|u_j)\}_{u_j \in U_{\mathbb{h}}}$, where $U_{\mathbb{h}}$ is the set of users in the household $\mathbb{h}$. Then, for each new event e, we obtain its representation as a feature vector $\hat{X}_e$, and identify the user who maximizes the PMF, that is, $u_j^*(e) = \arg\max_{u_j \in U_{\mathbb{h}}} P(\hat{X}_e|u_j)$. When more than one feature is used, we assume independence and use the joint probability function, i.e., the product of the features' PMFs.

The rest of the evaluated methods are Machine Learning (ML) algorithms able to deal with heterogeneous attributes. Specifically, we have restricted our study to the following methods: Bayesian Networks (BN), Logistic Regression (LR), and Decision Trees (DT). These methods provide a score distribution for a user (label) in a household (*class membership*). They provide a score $\{s(\hat{X}_e, u_j)\}_{u_j \in U_{\mathbb{h}}}$ based on different statistics from the training data, and select the users with highest scores.

3.2 Local Classification Methods

The methods introduced in the previous section use the same fixed set of features for every household in the system.

It may be the case, however, that members of a given household are particularly better discriminated by using a specific subset of features. For instance, let us consider a household where members always interact with the system at the same period of time (e.g. 8-10 pm) but in non-consecutive days (e.g. workdays vs. weekend), and compare it with other in which some of its members interact always after the others (e.g. children before 8 pm vs. parents after 9 pm). Although global methods may obtain accurate predictions for some non complex households, they may be too generic and thus could lead to worse performance in a test set.

In order to deal with households in which temporal habits can be better discriminated by specific subsets of features, in this section, we present local methods that select the best subset of attributes for each household. We explore two local methods, namely an attribute selection strategy from the ML field, and an ad-hoc approach from the Probability Theory, which is based on measuring distances between feature distributions. More specifically, the attribute selection method performs a search in order to select the subset of features that better correlate with the class (user label). For the ad-hoc approach, we use the KLD introduced before to measure the distance between distributions. The rationale behind this is that the closer two feature distributions are (in the context of a household), the less likely such features would discriminate the household members. Furthermore, it is possible for this method to define a "default feature" or even a ranking of features, if necessary.

4. EXPERIMENTS

In this section, we report and discuss results obtained in experiments we conducted to evaluate the discrimination power of the time features and the local methods presented in Section 3. We also evaluated two baselines for comparison purposes, namely a *Random* classifier, and a *Frequency*-based classifier that for a given test event, selects the household member who has the largest number of previous events in the training set, and no rating for the event's item.

We use a real movie rating dataset made publicly available by MoviePilot for the 2011 edition of the CAMRa challenge [6]. This dataset contains a training set of 4,536,891 timestamped ratings from 171,670 users on 23,974 movies, in the timespan from July 11, 2009 to July 12, 2010. The dataset also contains two test sets, intended for different challenge tracks: Test set #1, with 4482 ratings from 594 users on 811 items in the timespan from July 15, 2009 to July 10, 2010, and Test set #2, with 5450 timestamped ratings from 592 users on 1706 items in the timespan from July 13, 2009 to July 11, 2010. Test set #1 was intended for the rating prediction task, whereas Test set #2 was intended for the household member identification task.

We computed the accuracy of the evaluated methods in terms of the correct classification rate by household ($acc_{\mathbb{H}}$), i.e., the number of correct active member predictions divided by the total number of predictions, averaged by household, as proposed by CAMRa organisers. Formally, let $\mathbb{H}$ be the entire set of households in the dataset, and let $f(\cdot)$ be a method under evaluation. The metric $acc_{\mathbb{H}}$ is expressed as follows:

$$acc_{\mathbb{H}} = \frac{1}{|\mathbb{H}|} \sum_{\mathbb{h} \in \mathbb{H}} \frac{1}{|\mathbb{h}|} \sum_{(e_i, u_i) \in \mathbb{h}} L(u_i, f(e_i))$$

where $f(e_i) = \hat{u}$ is the user predicted by $f(\cdot)$ as associated to e_i, $L(u, \hat{u}) = 1$ if $u = \hat{u}$ and 0 otherwise, and (e_i, u_i) are the pairs of events and users in the test set of household $\mathbb{h}$.

By measuring the $acc_{\mathbb{H}}$ we empirically test the discrimination power of the time features when used by different methods (Section 4.1), and evaluate whether local methods outperform global methods (Section 4.2).

4.1 Discrimination Capabilities of Time Features

We first study whether time features alone are a valuable source of information that can be used to properly discriminate users in the identification of active household members. Table 2 shows the $acc_{\mathbb{H}}$ values obtained by the A priori method when using each of the proposed time features (Table 1), and using different combinations of such features. Note that in the table, the diagonal cells contain the $acc_{\mathbb{H}}$ values obtained from the use of a single feature, and the remainder cells contain the $acc_{\mathbb{H}}$ values obtained from the use of feature combinations.

We observe that the best single performing features are D, W and H, which is in accordance with the KLD-based feature ranking reported in Table 1, confirming the predictive power of KLD. In cases where two features were used, combinations including any of D, W and H features obtained better results. Furthermore, we evaluated all the possible combinations of features, and found that combinations including D, W and H achieved the best results. Table 2 shows results for some of those combinations. In particular, the best $acc_{\mathbb{H}}$ value of 0.9737 was achieved by combining the features D, W, H and Q_h.

We also evaluated BN, LR and DT Machine Learning methods. We used Weka[3] implementations of BN, LR and J48 DT algorithms, with default parameter values. Their accuracy values are shown in Table 3 for various combinations of time features. We observe that, in general, these methods outperform the A priori model for a small margin. We also note that as more features are used, the higher accuracy is obtained, although combining only D and H features achieve high accuracy values as well. The highest accuracy was obtained by the DT method when using all the considered time features. From these results we conclude that the identification of active household members can be effectively addressed by exploiting only timestamp information, regardless of the classification method used.

4.2 Classification Accuracy of Global and Local Methods

In this section we report the accuracy values of the local classification methods presented in Section 3.2, and compare them with those of the best performing global classification methods, and those of the proposed baselines. The comparison is conducted on the two available CAMRa test sets, aiming to check the generality of the proposed methods.

Table 4 shows the obtained results. Random and frequency-based baselines had a poor performance on Test set #1, and a better performance on Test set #2. This may be due to the differences on the rating data distributions in the test sets, which were built with distinct purposes. We observed that in Test set #1, every test item assigned to a household had not been previously rated by a member of the household. This fact turns the frequency-based classifier into a random classifier, since it is not able to decrease its uncertainty by getting rid of some of the users in the household (who previously rated the test event's item).

Regarding local approaches, we use an attribute selection method based on symmetrical uncertainty (an entropy based measure provided in the Weka toolkit) for the BN, LR and DT models, and KLD for the A priori model. The results reported in Table 4 show that the local methods are able to

[3]Available at `http://www.cs.waikato.ac.nz/ml/weka/`

Table 2: Correct classification rates obtained by the A priori method using different time feature combinations on Test set #2. Darker grey cells indicate worse values of the metric. Best value in bold.

	D	W	H	P_d	P_w	M	M_h	Q_h	M_y	DW	DH	DWH
D	0.9413											
W	0.9426	0.9310										
H	0.9727	0.9652	0.9457							0.9720		
P_d	0.9557	0.9467	0.9391	0.8260						0.9564	0.9720	0.9718
P_w	0.9430	0.9298	0.9531	0.8885	0.7991					0.9422	0.9721	0.9716
M	0.9553	0.9435	0.9402	0.8544	0.8614	0.7832				0.9554	0.9718	0.9720
M_h	0.9509	0.9424	0.9511	0.8944	0.8942	0.8793	0.8396			0.9514	0.9670	0.9670
Q_h	0.9517	0.9409	0.9532	0.8786	0.8770	0.8642	0.8404	0.8081		0.9511	0.9729	**0.9737**
M_y	0.9420	0.9372	0.9538	0.8472	0.8332	0.8077	0.8657	0.8351	0.7190	0.9430	0.9732	0.9722

Table 3: Correct classification rates of Machine Learning methods using different time features on Test set #2. Best value in bold.

	BN	LR	DT	A priori
D	0.9538	0.9515	0.9472	0.9413
W	0.9438	0.9405	0.9435	0.9310
H	0.9442	0.9432	0.9459	0.9457
TW	0.9484	0.9564	0.9470	0.9426
TH	0.9740	0.9769	0.9709	0.9727
WH	0.9690	0.9701	0.9750	0.9652
TWH	0.9744	0.9759	0.9752	0.9720
All	0.9722	0.9785	**0.9787**	0.9663

improve results only in some cases. Tests with some other attribute selection methods (not presented herein due to space limitations) indicate that the choice of the best attribute selection method depends on the used classifier. We note that the tendency of results of the local method is similar on the two test sets, although better results were obtained on Test set #2. The best result of the A priori model was obtained with the combination of DH features on both test sets, while the best result among ML models was obtained by LR using all features for Test set #1. On Test set #2, the best value were obtained by both LR using feature selection and DT using all features.

All these results show that correct classification rate is prone to minor differences depending on the utilized household member identification methods. The local methods not only obtain improvements in certain cases, but also have the benefit that they are more efficient than global approaches, since they require less attributes.

In any case, the use of adequate time features brings the most significant improvements, achieving much higher accuracy values than the random and frequency based classifiers.

5. CONCLUSIONS AND FUTURE WORK

In this paper we have presented and evaluated a number of methods to effectively identify which member in a household is currently interacting with an online recommender system at a particular time, by only exploiting knowledge about past user interactions with the system. We focused our study on analyzing existing differences in temporal rating habits, described in terms of various time features. These features were used to discriminate between users in a household by means of a classification algorithm. We identified the best performing time features for user discrimination purposes, and exploited such features with methods that address the identification of active household members globally –by using the same set of features for all the households– and locally –by dynamically selecting the most relevant features for each household.

The results obtained in our evaluations showed that simple algorithms are able to achieve high accuracy values when certain time features are used. They showed that isolated time features are valuable sources of information for discriminating users in a household (**RQ1**), and that there exist minor differences on accuracy values achieved by global and local methods (**RQ2**). Based on these results, we conclude that a simple and powerful approach to address the identification of active household members is to exploit the highest discriminant time features by a classifier with no local adaption for each household.

Table 4: Correct classification rates of the evaluated baseline, global and local methods on Tests sets #1 and #2. Best values in bold.

Model	Test set #1	Test set #2
Random	0.4988	0.4890
Frequency	0.4906	0.8100
A priori-All Features	0.9384	0.9663
A priori-DH	0.9504	0.9727
BN-All Features	0.9482	0.9722
LR-All Features	**0.9552**	0.9785
DT-All Features	0.9528	**0.9787**
A priori-KLD	0.9366	0.9667
BN-Attribute Selection	0.9476	0.9725
LR-Attribute Selection	0.9539	**0.9787**
DT-Attribute Selection	0.9517	0.9786

Next steps in our research will consider to exploit the discrimination power of time features for identifying underlying common preferences in scenarios where individual household members are not known –the common setting in current online services. Extracting such preferences would be equivalent to identify latent user profiles[3], i.e., specialized profiles related to individual users' interests. Tackling appropriately this latter task may enable services based on household accounts to provide truly personalized recommendations.

6. ACKNOWLEDGMENTS

This work was supported by the Spanish Government (TIN2011-28538-C02-01). The authors thank Centro de Computación Científica at UAM for its technical support.

7. REFERENCES

[1] P. G. Campos, F. Díez and A. Bellogín. Temporal Rating Habits: A Valuable Tool for Rater Differentiation. In *CAMRa*. ACM, 2011.

[2] J. Bento, N. Fawaz, A. Montanari, and S. Ioannidis. Identifying users from their rating patterns. In *CAMRa*. ACM, 2011.

[3] Y. Kabutoya, T. Iwata, and K. Fujimura. Modeling Multiple Users' Purchase over a Single Account for Collaborative Filtering. In *WISE*. Springer, 2010.

[4] Y. Koren. Collaborative filtering with temporal dynamics. In *KDD*. ACM, 2009.

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

[6] A. Said, S. Berkovsky, E. W. D. Luca, and J. Hermanns. Challenge on context-aware movie recommendation: CAMRa2011. In *RecSys*. ACM, 2011.

[7] Y. Shi, M. Larson, and A. Hanjalic. Mining relational context-aware graph for rater identification. In *CAMRa*. ACM, 2011.

[8] M. Zhang and N. Hurley. Avoiding monotony: improving the diversity of recommendation lists. In *RecSys*. ACM, 2008.