Finding Days-of-week Representation for Intelligent Machine Usage Profiling

Chiming Chang, Paul-Armand Verhaegen, and Joost R. Duflou
Centre for Industrial Management, Department of Mechanical Engineering, KU Leuven, Belgium
{Chiming.Chang, PaulArmand.Verhaegen}@cib.kuleuven.be, Joost.Duflou@mech.kuleuven.be

Madalina M. Drugan and Ann Nowe
Computational Modelling Lab, Artificial Intelligence, Vrije Universiteit Brussels, Belgium
{mdrugan, ann.nowe}@vub.ac.be

***Abstract*—Usage profiles of a smart appliance predict how the machine is expected to interact with its users according to its usage history, but, the problem of building usage profiles has been scarcely discussed in the literature. In this paper, we discuss general aspects of generating usage profiles and propose a daily pattern based probability model for usage profiling. We show how the probability models can be learned with Bayesian network classifiers and we highlight the importance of finding the optimal days-of-the-week representation. An algorithm using the conditional log-likelihood minimum description length (CMDL) and hierarchical clustering is designed to find the representation. The learned model is then used in a Bayesian network classifier setting to predict usage profiles. The methodology is tested on areal-life dataset of office printers in a campus environment.**

***Index Terms*—usage profile, MDL, Bayesian network classifier, days-of-the-week, smart home**

I. Introduction

The research of smart home and smart appliances design pictures a future where intelligent machines understand the user demands and provide the expected services accordingly. There are several advantages of a smart system. Energy efficiency can be achieved by switching off the power of appliances, lights and heating, when they are not requested by users. User comfort can be achieved by reducing the wait time of a service, by preheating a room right on time before users walking in, or by providing user-dependent services. Moreover, routine and repetitive human-machine interactions, e.g. turning on/off the lights and electronic switches, can be avoided and the device damage due to excessive operation reduced. Improved safety can be achieved by anomaly detection of appliance usage, e.g. a stove left on for a long time.

Usage profiles of smart appliances are estimations of how the machines are expected to interact with their users. This information tells us when, by whom and/or under what situations the machines are likely to be used. Building usage profiles of smart appliances is an attempt to find patterns and learn the user behavior from historical usage data. In this paradigm, we assume that the previous experiences can predict future situations. We consider as appliances all home/office devices, such as coffee makers, electric stoves, printers, etc., or a service in a smart environment, such as the heating and light of a room. Throughout this paper, the phrases 'machine' and 'appliance' are used interchangeably.

The problem of usage profiling is considered here a part of the smart home architecture. Typically, a smart home consists of sensors and smart appliances that are linked to a local network and a central server that stores the historical usage (or energy consumption data) and sensor readings. The central server has also the ability to remotely control the appliances. Ideally, an intelligent engine on the server analyzes the data, finds out the demands and makes decisions/policies for devices. Thus, usage profiles are an important building block in the knowledge bank of designing independent intelligent machines. In such applications, machines have abilities to record their usage history and dynamically adjust their behavior according to the expected demand.

Research a smart home and smart office can be categorized in different subtopics. MavHome [1], [2] presents the system architecture of a smart home and different algorithms on location and inhabitants action prediction. Context awareness study tries to model human mobility and daily activities using sensor data and home appliances usage records [3], [4]. Short term load forecasting (STLF) aims at power grid level demand prediction using different machine learning skills [5], [6].

Generating usage profiles for a smart appliance, on the contrary, is less addressed until recently. [7] uses Bayesian network to predict the usage of a single smart home appliance. Kaustav et al. [8], [9] propose a generic model using a knowledge driven approach to forecast the appliance usage. In [10], Chen et al. presents the daily behavior-based usage pattern algorithm to extract the usage pattern of the smart home appliances based on the hierarchical clustering. There is, however, not a generic methodology for usage profiling of a single smart appliance yet.

Manuscript received May 24, 2013; revised August 17, 2013.

doi: 10.12720/jiii.1.3.148-154

In this paper we discuss different aspects of building generic usage profiles for a smart appliance. We proposed a daily pattern based probability model and show how to learn the model from historical usage data using Bayesian network classifiers. An algorithm that finds the optimal days of the week representation is proposed. Bayesian network classifiers using this representation have better probability estimation compared to the ones using the conventional seven values representation corresponding to the different days of the week.

The remainder of this paper is organized as follows. Section II discusses general aspects of building usage profiles; the nature of the problem and their difficulties. In Section III, a daily pattern based probability model for usage profiling is proposed. We propose an algorithm to find the optimal days-of-the-week representation and show the corresponding experimental results in Section IV. The final section draws a short conclusion and sketches future research directions.

II. Generating Usage Profiles

A. Purpose

The purpose of building usage profiles is to learn a usage model from historical data. With this model, the conditional probability of machine usage under given circumstances, temporal attributes and environmental factors, can be estimated. An intelligent controller can dynamically determine the machine behavior based on this information and other constraints, such as user comfort or an energy consumption goal.

Although many classification and clustering methods have been tested to generate the usage models for smart appliances [7]-[9], it is important to address the difference between building usage profiles and the typical classification problems. For a prototypical classification problem the objective is to find a model which can correctly predict the class value of the highest number of test samples. For usage profiling, it is crucial that the model also provides precise probability estimation for each class value. Thus, the precision of probability estimation is more important than the classification accuracy, while a 10 or 40 percent chance of usage would typically be classified as ”no use”. From a usage point of view both are very different, as this may lead to different decisions or control policies of the intelligent controller. Second, the classification problems normally deal with a dataset consisting of a number of independent instances. Historical usage data, on the contrary, are time series. It is a straightforward approach to divide the usage history into evenly spaced time frames and use them as instances. The dependency between time frames are here neglected.

B. Properties of Usage Profiles

We discuss several properties of usage profiles below.

1) Temporal characteristics: The usage of a home/office appliance has a temporal nature and is often subject to schedules of its users. Machines used in an office environment show distinct daily (on/off work) and weekly (weekday/weekend) patterns that mainly depend on the presence of the users. House appliances are used mainly at nights and during weekends. Late meetings at the office routine or family vacation are exceptions that generate irregularities into an household schedule. Generally speaking, temporal factors -days of the week and hours of the day- provide the most information about the machine usage. Short term dynamics (previous hours or past few days) and season factors (the same month last year) may also be a factor in predicting the usage, but are often ad-hoc and depend on the application. In general, the machine usage can be viewed as a time series with strong daily and weekly patterns, but are often interrupted by random and unpredictable events.

2) Probabilistic nature: The usage of an appliance is probabilistic in nature. Identical printers installed in different locations of an office building have different usage patterns depending on their users. Printers shared by many users have a higher utility rate while other printers are rarely used, even during office hours. The machine usage usually takes the format of categorical values, such as used/no-used, off/on/standby and etc. The short term load forecasting (STLF) compares the daily energy consumption of a building where an energy consumption reading is a continuous variable resulting from aggregated behavior. STLF is useful when it shows clearer patterns and the energy consumption is quasi-stable. Thus, we cannot use this model to compare the day to day usage of an appliance because of the large variances in patterns. For usage profiling, the challenge is finding out the probability of each categorical value. This is discussed in the next section.

3) Small datasets: From the statistical point of view, the more data samples we have, the better estimation of probability we get. However, we face the problem of small datasets while trying to build a usage profile. One reason is that we want to obtain a usage profile with a high credibility as soon as possible so the automation policy can be deployed in an early stage of system use. Even if we have the usage history for two or three years, there is the problem that human schedules or user habits may change over time. We assume that the ’recent’ data are more informative than data, and, thus, we build the usage profiles from a limited number of samples. In conclusion, the difficulty in usage profiling is that the usage probabilities have to be estimated based on limited data in which the patterns are often interrupted.

III. A Generic Model for Usage Profiling

We introduce a generic model for usage profiling of smart appliances.

A. The Model

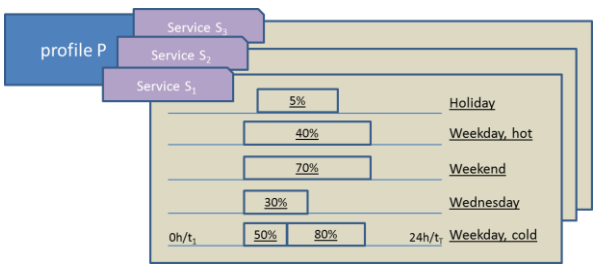

Figure 1. A daily pattern based probability model for usage profiling.

The proposed model defines a machine as a set of services provided. For each service, several day-types are identified and the probabilities of service requested for each time frame within a day are listed. An example is given in Fig. 1.

Definition 1: Let $S = \{s_1, s_2, \ldots s_N\}$ be the set of services of machine M. Usage profile U of machine M is defined over S. For each s_i, $U(s_i) = \{(\hat{P}_j, C_j) : j \in 1, \ldots K\}$, where j is the index of K identified day clusters for service s_i. $\hat{P}_j = \langle p_1, p_2, .. p_T \rangle$ is the probability estimation of usage for s_i on T equally spaced time frames of a day. C_j is the characteristic description of the day cluster.

The proposed model defines the structure of the information that we try to obtain from the historical data. The *service set* defines the services a machine can provide and is relevant to the controller. The most fundamental and important service set is $\langle on, off \rangle$ of machine power. We can refine this set in order to model better the states of machine. For example, some machines have a "sleep" or "stand-by" mode and some have a "pre-heating" mode between on and off. Then, the service set has five states $\langle on, off, sleep, stand\text{-}by, pre\text{-}heating \rangle$.

For a coffee maker, for example, this model can be even more elaborated including a certain type of coffee or the requests from a particular user.

For each service, a profile consisting of a number of identified day-types is built. Each day-type is composed of probability estimation for every time frame of a day ($\hat{P}$) and a characteristic description (C) for the cluster. In this model, it is assumed that there are different day types of machine usage and there is an intrinsic probability distribution for each time frame of these day types. The intrinsic probabilities are actually reflections of the *contexts* of machine usage, such as office hours, after work, early in the morning, etc. We assume that *one day* is the natural segmentation of time for pattern searching and clustering. We claim this is appropriate because machine usages, for most cases, are subjected to human schedules and activities, which are normally day-based.

Here, the challenge in analyzing the data is two-fold. We need to identify the day types by their difference in the probability of usage. On the other hand, a good identification of day types helps to properly estimate the probability distribution. For example, from the historical data, if there are 5 day instances that the machine is used and 5 day instances the machine is not used, we may conclude that the machine has a 50 percent probability of usage. However, if we consider that 4 of the 5 day instances during which the machine is not used are holidays, it seems better to separate the holiday and non-holiday groups. However, if there are 2 used holidays and 2 not-used holidays, then the holiday factor is not so important.

B. Generating the Model

We show how the models of printers used in an office environment can be generated using Bayesian network classifiers. We highlight the problem of finding an optimal days-of-the-week representation which arises in the process.

1) The dataset: The dataset used is a data log of printer usage in an office building of KU Leuven. The data log consists of the usage of 53 printers over a 3 years period, from 2009 to 2012. However, some of the printers have only a short history of usage and some are used sparsely. After removing these printers, 21 printers were left for analysis.

Usage records of printers take the form of $\langle time, printer, user, pages, copies \rangle$ tuples. The unit of time frame is selected to be 1 hour. The service is defined to be *the existence of any print jobs in the time frame*. In other words, we are asking the question what is the probability of a printer receiving a printing request given a certain hour.

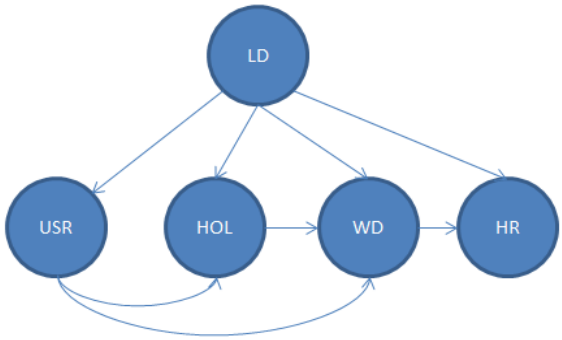

Figure 2. An example of a Bayesian network classifier to learn the usage model

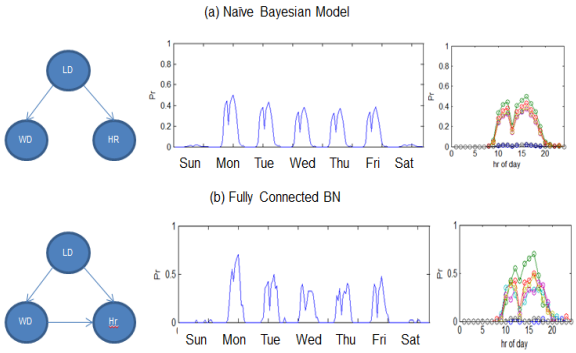

Figure 3. Two types of Bayesian network classifiers with respective day patterns.

2) Building a Bayesian network classifier: In order to estimate the hourly probability of usage, we use Bayesian network classifiers (BNC) [11]. Fig. 2 gives an example BNC that can be used to learn the usage profiles where the class variable LD (loading) is determined by four attributes $USR, HOL, WD, and\ HR$, representing the effect of users, holiday, days of the week, and hour of the day. To represent the days of week attribute, it seems trivial to use a seven values mapping like $WD \in \{1,2, \ldots 7\}$ where 1 for Sunday, 2 for Tuesday, etc. However, we point out here that for usage profiling this may not be the best representation.

We consider only the effect of hours of the day and days of the week. The record based usage data is then

transformed into a dataset consisting of hourly samples. Each sample is a tuple of $\langle \mathrm{WD}, \mathrm{HR}, \mathrm{LD} \rangle$. The attribute $WD \in \{1,2, \ldots 7\}$ represents days of the week, from Sunday to Saturday. $HR \in \{1,2, \ldots 24\}$ Indexes hours of the day. Variable $LD \in \{0,1\}$ indicates the existence of print jobs in the hour. We define LD as class variable and WD, HR as attributes. In this simplified case, only two models are possible, a naive Bayesian model and a fully connected Bayesian network.

Fig. 3 shows the networks and the probability estimation learned for each day of the week for printer No. 38. The two models start from different assumptions and estimate the probability accordingly. The naive Bayesian model assumes that attributes LD and WD are independent from each other. Thus, it takes the empirical probability distribution of all days and weighs it with day-of-week distributions. This is clarified in the equation below.

$$\mathrm{p}(\mathrm{LD}|\mathrm{WD}, \mathrm{HR}) = \frac{\mathrm{p}(\mathrm{WD}|\mathrm{LD})\mathrm{p}(\mathrm{HR}|\mathrm{LD})\mathrm{p}(\mathrm{LD})}{\mathrm{p}(\mathrm{WD},\mathrm{HR})} \quad (1)$$

The problem for the naive Bayesian model is that all days of the week share a similar daily usage pattern with only a difference in magnitude. This is obviously contradicting our daily experience. Another model is obtained by adding an arc, representing a conditional dependence, between WD and HR (the fully connected model). This model assumes that WD and HR are conditionally dependent, and considers the days-of-week factor in the conditional probability table of node HR. In this model, each day-of-week is represented by a distinct usage pattern which is in fact the empirical probability of each day-of-week in the database. The model, however, is in contradiction with our daily experience from another perspective. The probability distributions of *Tuesday, Wednesday, Thursday and Friday* are similar to each other. These days are different day types with minor difference in intrinsic probabilities, it seems more reasonable to assume that these days belong to a certain day type, normal working day, and the discrepancies of empirical probabilities are merely a sampling bias.

For a machine installed in an office environment, according to our experiments, it is better in terms of performance to assign the representation $WD \in \{weekday, weekend\}$ than to assign$WD \in \{Sunday, Mondday, .. Saturda$. However, the printers can be situated in a home or other environment where the representation does not apply. There may also be exceptions that some offices, for example, only work half-days on Friday and maybe some have a part-time work schedule. In such case, a more refined representation is required. We want to find the best days-of-week representation from a machine's usage history automatically.

IV. Finding Optimal Days-of-the-Weekrepresentation

A. The Problem

We now formulate the problem of optimal days-of-the-week representation. Given the usage history of a machine and to consider only hours-of-the-day and days-of-the-week factor, we investigate how find the optimal representation for days-of the-week attribute which better represents the machine usage patterns.

We use the serial notation for days-of-the-week representation. The representation $\langle 1234567 \rangle$ assigns Sunday to 1 , Monday to 2 and etc. Representation $\langle 1222221 \rangle$ assigns a weekday/weekend schedule. These numbers are used to show which day type a day-of-week belongs to and the first position is always Sunday. We assume that there is a hidden weekly structure for machine usage depending on the environment and user behavior. Representation $\langle 1234567 \rangle$ means each day of the week has its own usage pattern and representation $\langle 1111111 \rangle$ indicates one daily pattern for all days. The goal is to find this 'true' representation from the historical data and by applying this information we can build a model with better probability estimation.

B. The Algorithm

We present the pseudo-code for Algorithm 1 for the problem of optimal days-of-the-week representation.

Algorithm 1 Finding optimal days-of-the-week representation

Input: historical usage data, $D = \{u_1, u_2, ..u_S\}$, where $u_i \in WD \times HR \times LD$. WD is defined by a mapping $< Sun, Mon, .., Sat > \rightarrow < 1, 2, .., 7 >$.

Output: a new mapping M for the days-of-the-week factor

1: Find all combination of two pairs $(i, j) \in Val(WD)$, $i \neq j$.
2: $\forall i, j$ pair, create new attribute WD_{ij}, where $WD_{ij} = WD$, except that i, j take same value assignment which is different from other values.
3: For each new attribute WD_{ij}, calculate $I_P(WD_{ij}; HR|LD)$
4: Consider a hierarchical clustering problem having $N = |Val(WD)|$ instances. Define dissimilarity $dis(i, j) = I_P(WD; HR|LD) - I_P(WD_{ij}; HR|LD)$.
5: Generate the dendrogram.
6: For each possible number of clusters, (1 to N), calculate the weighted $CMDL$.
7: best cut $\leftarrow$ minimum CMDL.
8: Generate mapping M from the best cut.

We use the reduction of conditional mutual information as the dissimilarity measure between days of the week in step 1 to 3. The conditional mutual information $IP\,(WD; HR|LD)$ is calculated using Equation 2. This metric represents the dependency between WD and HR on condition of LD [11]. For the seven day-types define by WD, Sunday to Saturday, if the variance of distribution $\hat{P}(LD|HR)$ is large then the value is large, and vice versa. We can combine any two day-types and create a new attribute, i.e. WD_{17} , for $\langle 1234561 \rangle$. The difference between $I_p(WD; HR|LD)$ and $I_p(WD_{17}; HR|LD)$ reflects the loss of information for new attribute WD_{17} in which Sundays and Saturdays are

regarded as the same group and hence it cannot be identified. The more different the distributions $\hat{P}(LD|HR)$ for Sunday and Saturday are, the larger the difference. It is 0 when they are identical. The differences for all value pairs of WD are calculated and used as the dissimilarity measure.

$$\mathrm{I_P(WD;HR|LD)} = \sum_{wd,hr,ld} P(wd,hr,ld) \log \frac{\mathrm{P(wd,hr|ld)}}{\mathrm{P(wd|ld)P(hr|ld)}} \quad (2)$$

A dendrogram is then built by hierarchical clustering (step 4). In general, we can cut any level of the dendrogram to get a new representation. The generated representation will have 7 to 1 clusters (day types) respectively. The more clusters kept, the less information loss there is. It is crucial to point out that it is not always optimal to keep all the information content, because some information contents simply originates from the difference between the empirical and the intrinsic probability distribution and thus is redundant. The idea is to merge as many values as possible but to stop when the information loss is too large.

To determine the best cut level for hierarchical clustering, we adapt the scoring function of the *conditional minimum description length* (CMDL)[12]. The *minimum description length* (MDL) score is a widely used quality measure for comparing Bayesian networks. The main idea is to find an optimal model which best describes the data (maximum likelihood) with minimum network complexity. Equation 3 shows the standard form for the MDL score. In this two part code, the left-hand part represents the complexity of Bayesian network B while the right-hand part is the log-likelihood function for data Dgiven$B$. For Bayesian network classifiers, a natural extension is to use CMDL, in which the log-likelihood function is replaced by the conditional log-likelihood (Equation 4). One major difficulty for using the CMDL score for Bayesian network classifier searching is that the joint probability of attributes does not factorize over the network. This, however, is not an obstacle here. In the proposed method, we simply use the empirical conditional probability to calculate the CMDL score. For each cut level, i.e. 1 to 7 clusters, we create a new attribute WD^* which is a mapping of the new representation for the days of the week from the clustering result. The attribute WD^* is then used to calculate the conditional log-likelihood and CMDL score (Equation 4). The cut level with minimum CMDL score is then the best representation for days of the week attribute for the given dataset.

$$\mathrm{MDL(B|D)} = \frac{\log N}{2}|B| - LL(B|D) \quad (3)$$

$$\mathrm{CMDL(B|D)} = \alpha . \frac{\log N}{2}|B| - CLL(B|D)$$

$$\mathrm{CLL(B|D)} = \sum_{i=1}^{N} \log(P_B(ld_i|wd_i^*, hr_i)) \quad (4)$$

We also introduce the weight factor α in Equation 4 to control the level of complexity for the output representation. The value is pre-selected and does not change over different cases.

C. Experimental Results

We use Algorithm1 for all printers in our dataset using 40 weeks historical data and tried to find the optimal days of the week representation for each printer. Fig. 4 shows the process of applying the method on printer No. 38 with a usage history of 40 weeks. Figure 4(a) shows the dendrogram of hierarchical clustering results using reduced conditional mutual information as a similarity measure. For printer 38, there is a distinct weekday/weekday pattern, being heavily used on Mondays compared with the other days of the week. The tendency can also be observed from average days-of-the-week pattern in Fig. 4(c).

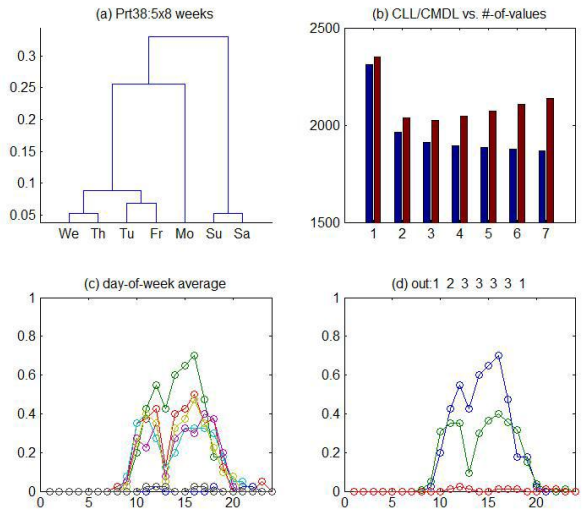

Figure 4. An illustrative example of applying Algorithm 1 on usage history of 40 weeks for Printer. No 38.

Note that the usage patterns vary from printer to printer. Some have consistent weekday patterns. Some have one or two weekdays different from others. This leads to different clustering results and produces different optimal representations.

We choose the cut level at 3 clusters for printer No. 38 because the CMDL score in Fig. 4(b) shows the minimum. The resulting clusters - Cluster1: $\langle Sunday, Saturday \rangle$, Cluster2: $\langle Monday \rangle$, Cluster3: $\langle Tuesday, Wednesday, Thursday, Friday \rangle$ - can then be interpreted as a new representation $\langle 1233331 \rangle$. A usage model built upon this new days-of-week representation consists of 3 daily patterns which are characterized by their days-of-the-week labels (Fig. 4(d)). The benefit of using 3 instead of 7 clusters is that there are more day samples for each cluster, except the Monday cluster. According to the law of large numbers, the empirical probability is therefore a better estimation to the intrinsic probability. This is especially advantageous because the number of samples is usually limited and the results are therefore sensitive to data noise when performing usage profiling.

A stratified 5-fold cross validation is used to check the performance of the proposed method. For each printer, two predictive models built with fully connected Bayesian network are compared. The first one uses the $\langle 1234567 \rangle$representationfor days of the week attribute

and the second one uses theoptimal days-of-the-week representation found by Algorithm1.Table I shows the validation results and optimal days-of-the-week representation found for 8 printers. Each row represents one printer dataset. The classification accuracy (CA) is similar between the two models. This is reasonable because days of the week patterns are merged only when they have similar probability estimations for each time frame. The conditional log-likelihood (CLL) measure for models using optimal days-of-the-week representation is in general better for all printers. This corresponds to our belief that a machine-usage dependent representation other than ⟨1234567⟩ can be used for the days-of-the-week attribute to help estimate the usage probability better.

The rest printers in the database show consistent results and most of which have a ⟨1222221⟩ optimal representation.

TABLE I. TABLE OF VALIDATION FOR 8 PRINTER DATASET

No.	optimal representation	5 fold CA		5 fold CLL	
		⟨12..7⟩	optimal	⟨12..7⟩	optimal
38	⟨1233331⟩	0.8677	0.8680	-1513.5	-1460.3
37	⟨1222221⟩	0.8097	0.8095	-2351.1	-2208.0
11	⟨1232321⟩	0.8711	0.8711	-1359.1	-1341.7
25	⟨1222261⟩	0.8599	0.8594	-1512.8	-1499.0
30	⟨1224561⟩	0.8523	0.8525	-1625.1	-1599.5
26	⟨1222221⟩	0.8204	0.8205	-2034.2	-1949.5
6	⟨1222221⟩	0.8168	0.8164	-2066.0	-1898.4
1	⟨1212331⟩	0.9039	0.9039	-1089.2	-1082.3

D. Discussion

In generating usage profiles for smart appliances, it seems trivial to use a seven-value representation for the days of the week factor. We claim that this is often not the best representation for weekly patterns. This representation assumes that each day of the week is unique and different from the others while neglecting the fact that some days of week may be characterized by similar behavior patterns. This results in suboptimal estimations of probabilities of usage. We suggest an approach to find out the weekly structure (a clustering for days of the week) first, and use the new representation to build the probability model.

V. CONCLUSION AND FUTURE WORK

In this paper, we discuss the purpose and properties of usage profiling. We propose a daily pattern based probability model for usage profiling and show how the model can be learned with Bayesian network classifiers. We highlight the problem of using the seven value representation for the days-of-the-week attribute and propose an algorithm to find the optimal representation. We claim that models using the optimal representations provide better probability estimations. We tested the methodology on a dataset of office printers and proved its feasibility.

In this preliminary research we assume each day of the week has a constant usage behavior and we cluster the similar days. This does not always correspond to our daily life reality. Public holidays and seasons are important factors in human behavior and, thus, in machine usage. Based on the methodology proposed, we aim to expand the model with these factors in order to generate a more realistic probability model.

REFERENCES

[1] S. Das, D. Cook, A. Battacharya, I. Heierman, E. O., and T. Y. Lin, "Therole of prediction algorithms in the mavhome smart home architecture," *Wireless Communications*, IEEE, vol. 9, no. 6, pp. 77–84, 2002.

[2] D. Cook, M. Youngblood, E. Heierman, K. Gopalratnam, S. Rao, A. Litvin, and F. Khawaja, "MavHome: An agent-based smart home," in *Proceedings First IEEE International Conference on Pervasive Computing and Communications*, 2003., pp. 521–524.

[3] V. Osmani, S. Balasubramaniam, and D. Botvich, "A bayesian networkand rule-base approach towards activity inference," in *Proc. Vehicular Technology Conference*, 2007, pp.254–258.

[4] A. Roy, S. Das, and K. Basu, "A predictive framework for locationawareresource management in smart homes," *IEEE Transactions on Mobile Computing*, vol. 6, no. 11, pp. 1270–1283, 2007.

[5] G. Chicco, R. Napoli, and F. Piglione, "Load pattern clustering for short-termload forecasting of anomalous days," in *Proc. IEEE Porto Power Tech* (Cat. No.01EX502), 2001, p. 6.

[6] H. Hippert, C. Pedreira, and R. Souza, "Neural networks for short-termload forecasting: A review and evaluation," *IEEE Transactions on Power Systems*, no. 1, pp. 44–55.

[7] L. Hawarah, S. Ploix, and M. Jacomino, "User behavior prediction inenergy consumption in housing using Bayesian networks," *Artificial Intelligence and Soft Computing*, pp. 372–379.

[8] K. Basu, M. Guillame-bert, H. Joumaa, S. Ploix, and J. Crowley, "Predicting home service demands from appliance usage data," in *Proc. International Conference on Information and Communication Technologies and Applications*, 2011.

[9] K. Basu, V. Debusschere, and S. Bacha, "Appliance usage prediction using a time series based classification approach," in *Proc. IECON 38th Annual Conference on IEEE Industrial Electronics Society*, 2012, pp. 1217–1222.

[10] Y. C. Chen, Y. L. Ko, and W. C. Peng, "An intelligent system for mining usage patterns from appliance data in smart home environment," in *Proc. Conference on Technologies and Applications of Artificial Intelligence*, 2012, pp. 319–322.

[11] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, pp. 131–163.

[12] M. M. Drugan and M. A. Wiering, "Feature selection for bayesiannetwork classifiers using the MDL-FS score," *Int. J. Approx. Reasoning*, vol. 51, no. 6, pp. 695–717, 2010.

Chiming Chang received the MS degree in electrical engineering from National Taiwan University in 2000 and the MS degree in artificial intelligence form the University of Leuven in 2010. He worked for the Media Tek, a Taiwan based IC design company, form 2000 to 2007. Chiming Chang joined the Centre for Industrial management research group of the University of Leuven where he currently performs a PhD study on the application of artificial intelligence to usage profile extraction for products.

Madalina M. Druganobtained a PhD (2006) from the Computer Science Department, University of Utrecht, The Netherlands. Her PhD thesis "Conditional log-likelihood MDL and Evolutionary MCMC" is researching (designing, analyzing, experimenting) various Machine Learning algorithms in fields like Bayesian Network classifiers, Feature Selection,

Evolutionary Computation, Markov chain Monte Carlo. As a postdoctoral fellow, she did research in algorithmic design for Bioinformatics, Multi-objective optimization, Meta-heuristics, Operational Research, and Evolutionary Computation. She currently works at Computational Modeling Lab, VrijeUniversiteit Brussels, Belgium.

Paul-Armand Verhaegen holds a master degree in applied science and engineering, specialisationelectrotechnical and computer science, a graduate in the complementary studies in business administration, and a MBA from the VrijeUniversiteitBrussel. He has founded and sold Stocks, a company specialised printer supplies. He has worked as an external consultant for 3E on green energy certificates, and as a consultant at Bureau van Dijk Management Consultants. He has worked at VrijeUniversiteitBrussel, ErasmushogeschoolBrussel and KU Leuven as a researcher and assistant. He has obtained a PhD in the domain of systematic innovation from KU Leuven, and is now a postdoctoral researcher at KU Leuven working on innovation and data mining.

Ann Nowe graduated from the University of Ghent in 1987, where she studied mathematics with optional courses in computer science. Then she became a research assistant at the University of Brussels where she finished her PhD in 1994 in collaboration with Queen Mary and Westfield College, University of London. The subject of her PhD is located in the intersection of Computer Science (A.I.), Control Theory (Fuzzy Control) and Mathematics (Numerical Analysis, Stochastic Approximation). After a period of 3 years as senior research assistant at the VUB, she became a Postdoctoral Fellow of the Fund for Scientific Research-Flanders (F.W.O.). Nowadays, she is a professor both in the Computer Science Department of the faculty of Sciences as in the Computer Science group of the Engineering Faculty.

Joost R. Duflou holds master degrees in architecturaland electro-mechanical engineering and a PhD inengineering from the KULeuven, Belgium. His principalresearch activities are situated in the field of designsupport methods and methodologies, with specialattention for Systematic Innovation, Ecodesign and LifeCycle Engineering. He is a member of CIRP and haspublished over 200 international publications.