PREDICTING HOME SERVICE DEMANDS FROM APPLIANCE USAGE DATA

Kaustav Basu*, Mathieu Guillame-Bert†, Hussein Joumaa*, Stephane Ploix* and James Crowley†
*G-SCOP lab
46, avenue Flix Viallet F-38031 Grenoble cedex 01
Email:Kaustav.Basu@ensimag.imag.fr, Hussein.Joumaa@imag.fr, Stephane.Ploix@inpg.fr
†INRIA Grenoble - Rhone-Alpes
655 avenue de l'Europe 38 334 Saint Ismier Cedex France
Email:Mathieu.Guillame-Bert@inrialpes.fr ,James.Crowley@inrialpes.fr

***Abstract*—Power management in homes and offices requires appliance usage prediction when the future user requests are not available. The randomness and uncertainties associated with an appliance usage make the prediction of appliance usage from energy consumption data a non-trivial task. A general model for prediction at the appliance level is still lacking. In this work, we propose to enrich learning algorithms with expert knowledge and propose a general model using a knowledge driven approach to forecast if a particular appliance will start at a given hour or not. The approach is both a knowledge driven and data driven one. The overall energy management for a house requires that the prediction is done for the next 24 hours in the future. The proposed model is tested over the Irise data and the results are compared with some trivial knowledge driven predictors.**

***Keywords*-Appliance Usage Prediction, Enriched Learning Algorithm, Energy Management in Homes, Data Mining.**

I. Introduction

Reducing housing energy costs is a major challenge of the 21^{st} century. In the near future, the main issue for civil engineering is the thermal insulation of buildings, but in the longer term, the issues are those of "renewable energy" (solar, wind, etc) and "smart buildings". Home automation system basically consists of household appliances linked via a communication network allowing interactions for control purposes [1]. Thanks to this network, a load management mechanism can be carried out: it is called *distributed control* in [2]. Load management allows inhabitants to adjust power consumption according to expected comfort, energy price variation and CO_2 equivalent emissions. A home energy management system able to determine the best energy assignment plan and a good compromise between energy production and energy consumption [3]. In this study, energy is restricted to the electricity consumption and production. [4], [3] present a three-layers (anticipative layer, reactive layer and device layer) household energy control system. This system is both able to satisfy the maximum available electrical power constraint and to maximize a ratio between user satisfaction and cost. The objective of the anticipative layer explained in [5] is to compute plans for production and consumption of services.

Uniqueness of housing systems involves a set of new issues in control system science: it is necessary to develop new tools [6], [7], [8] and algorithms [9], [10] for globally optimized power management of the home appliances, able to anticipate difficult situations but also able to take into account the actual housing system state and the occupant expectations.

Anticipating problematic situations require also prediction capabilities. Even if it is easier to predict overall consumption, it is important to be able to predict the consumption of each appliance because, regarding dynamic demand side management, it is also important to evaluate how much energy can be saved thanks to request to customers like unbalancing requests or energy price variations. The energy savings depend on appliances: some can be unbalanced, some can be postponed and some cannot be changed. The overall goal of the prediction is described in figure 1. It also includes an user interface where the user may provide his plans for the future. The proposed approach is restricted to the prediction of appliance usage using only appliance consumption data and time of the event.

The problem of appliance usage prediction through consumption data is new. [11] deals with the problem of the user behavior prediction in a home automation system using a Bayesian network for a single appliance but a general model for appliance prediction is still lacking. Short term load forecasting (STLF) at the grid level has been there for some time but at the appliance level, these techniques are yet to be tested. Though STLF uses regressive approaches whereas the proposed approach is based on classification but the strategies used in the domain of energy load prediction led to the choice of input to the predictor.
[12] does a study on the approaches used in load prediction. The approaches range from using methodologies such as similar day, expert knowledge and linear and non linear learning algorithms. [13], [14], [15], [16] gives details of implementation of neural networks in the domain of energy load forecasting and [17] proposed a SVM model to predict daily load demand for a month.

The objective of this work is to build an enriched learning

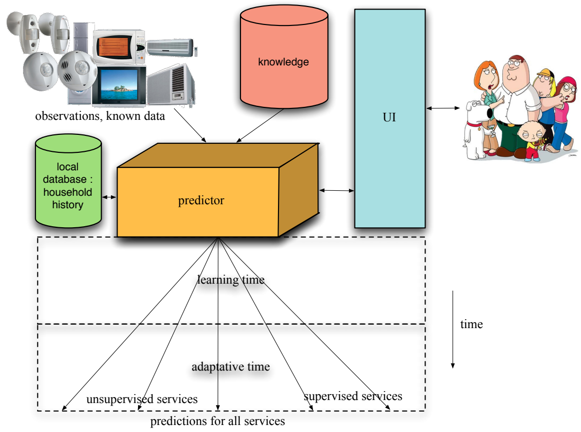

Figure 1. Goal of the prediction system

algorithm which takes knowledge into account and formalize it to statistically predict the user energetic service request for the next 24 hours. For the prediction of appliances from consumption data we first reduce the problem to a two class classification problem, i.e if an appliance is consuming at a particular hour or not. The model for the prediction is for each appliance in each house. The time space is sampled into 24 hours which aim's to predict the user appliance usage requirement for a particular hour. At each point of time the Prediction system will predict for the following 24 hours and then shift to the next hour and predict the following 24 hours.
In the approach an expert proposes certain knowledge based on his domain expertise and then to formalize and represent this knowledge. The knowledge representation is considered in an incremental manner and at every stage validate the knowledge in terms of accuracy of prediction. The organization of the paper is as follows, firstly, the Proposed model (section II) is discussed in details followed by choice of classifier and parameters (section III). The Oracle results as well as the overall results are presented in (section IV and V). Finally discussion about the results and the conclusion is drawn in (section VI and VII).

II. Proposed Model

The proposed model consists of enriched learning algorithm which proposes a general way to take expert knowledge into account. The proposed model divides the task into modules each of which has its own purpose. The general model is shown in figure 2 and in the following sub-sections each of the processing modules is discussed in details.

- Raw data contain
  - Energy consumption for an appliance .
  - Contextual information (Time, Date, Weather).
- Oracle is composed of statements leading to entities (factors) that may be taken into account.

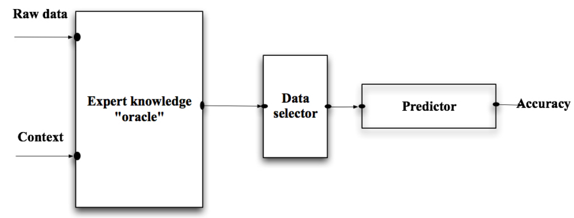

Figure 2. Schematic representation of the prediction system

- Data Selector is a processor which stores, selects and structures the data in order to present them to the classifiers.
- Predictor.

A. Database

A database is obtained from Residential Monitoring to Decrease Energy Use and Carbon Emissions in Europe (REMODECE) which is a European database on residential consumption, including Central and Eastern European Countries, as well as new European Countries (Bulgaria and Romania). This database stores the characterization of residential electricity consumption by end-user and by country. The IRISE project has been chosen from REMODECE which deals only with houses in France. Each database concerns one house. In such a database, information is recorded every 10 minutes for each appliance in house and over one year. This information represents the consumed energy by each service, its data and its time. Moreover, it is possible to know the number of people who live in each house. However, this data is not directly available. Let us notice that appliances are just involved in services: they are not central from the inhabitant point of view. Consequently, they are not explicitly modeled. The presence of the user is important but it is not predictable at the moment.

B. An expert system that generates knowledge : the Oracle

We define Oracle knowledge as statements leading to generated data (or factors) that may be taken into account. The Oracle receives the raw data from the database giving the consumption at an particular hour and the date, time and weather information at that hour. The Oracle proposes knowledge and then gives the necessary function which represents the data in a form interpretable by the Prediction system. The knowledge which are relevant for a particular appliance in a house might not be relevant for another house using the same appliance. So all knowledge proposed by the Oracle have to be validated and knowledge which doesn't increase or reduces the accuracy of prediction for a particular appliance have to be rejected. The part of validating and structuring of the Oracle output is done in the subsequent processing module as seen in figure II.

Statements and the functional representation inside the Oracle:

- Immediate past history of consumption is meaningful to appliance usage prediction.
- Hour of the day is meaningful to appliance usage prediction.
- Day of the week is meaningful to appliance usage prediction.
- Season of the year is meaningful to appliance usage prediction.
- Previous days same hour is meaningful to appliance usage prediction.

In the following sub-sections each of the proposed knowledge by the Oracle and their representation is looked in details.

1) Past Consumption History: The Oracle proposes that the past sequence of energy consumption prior to an event is meaningful in appliance usage prediction. It is represented mathematically followed by an illustration.

Mathematically, it is formalized by the following predicate function

Inputs: Consumption(*H*-1); Consumption(*H*-2);...; Consumption(*H*-*n*) ;

where,

n is the size of the past time history

H is the Current hour

Output : $\{0,1\}^n$

Here the output is a thresholded binary vector of size *n* which signifies if there is consumption at the hours prior to the event on not.

2) Hour of the Day: The Oracle proposes that the "time of the day" is meaningful to appliance usage prediction. In practice, by this knowledge the time space is discretized into 24 hour slots and the "actual time" the event occurs assumes priority. Firstly it is represented mathematically and then provided some illustrations to better understand the representation.

Predicate function HOD

Inputs : Current Hour in the day, X

$X \in \{(0-1),(1-2),(3-4),(5-6),...,(23-24)\}$ Hours

Output : $\{0,1\}^{24}$

This is an orthogonal representation of an hour rather than a numeric value.

Illustration :

If the Time of the day is 6.00 am,
instead of using the numeric value 6 we use "0, 0, 0, 0, 0, 1, 0,...,0" to represent the same.

So for a day example, the representation will be :

Hour 0	—	(1,0,0,0,0,...,0)
Hour 1	—	(0,1,0,0,0,...,0)
.	—	(0,0,1,0,0,...,0)
.	—	(0,0,0,1,0,...,0)
Hour 23	—	(0,0,0,0,0,...,1)

3) Day of the Week: The Oracle proposes that "Day of the week" is meaningful in appliance usage prediction. Similar to the way in II-B2 by taking this knowledge into account the whole week is discretized into 7 days. Instead of representing this with a numeric value, we use the orthogonal representation as in II-B2.

Predicate function DOW

Inputs : Current day of the week, X

$X \in \{Sunday, Monday, ..., Saturday\}$

Output : $\{0,1\}^7$

4) Season of the Year: Similar to the prior sub-sections the Oracle proposes that season of the year is meaningful in appliance usage prediction. There are appliances in houses which show distinctive different behavior depending on the season of the year. As like the prior representations an orthogonal representation is chosen over a numeric one.

Predicate Function SOY

Inputs : Current season of the year, X

where $X \in \{Spring, Summer, Autumn, Winter\}$

Output : $\{0,1\}^4$

So the season Oracle output will be:

5) Previous Days Same Hour: Here the Oracle proposes that what happens on previous days on the same hour is important in appliance prediction. So we look if there is consumption or not in the previous days for the same hour. The output is a vector of thresholded binary values of previous days at the same hour.

Predicate function

Inputs: consumption(*H*-24); Consumption(*H*-48);... Consumption(*H*-*n*);

where *n* is typically taken as 168 (one week)

H is the Current Day

Output : $\{0,1\}^7$

6) Oracle Output: The overall Oracle output after the representation of all the knowledges proposed by the Oracle is shown in table II-B6 where each row represent the proposed knowledge at a particular time in a incremental manner. The table is obtained by the incremental addition of knowledge proposed by the Oracle, so the knowledge proposed in section II-B1 to II-B5 are added incrementally in order. The Oracle has an available memory, thereby every hour in the history is represented by table II-B6.

C. Data Selector

The data selector is defined as an non-temporal matrix processor which stores, selects and structures the data for the predictor. This matrix is the input to the predictor. The

Knowledge	***H***
Consumption(H-1)	0/1
Consumption(H-2)	0/1
...	0/1
Consumption(H-n)	0/1
Hour of the day(0-1)	0/1
Hour of the day(1-2)	0/1
...	0/1
Hour in the day(23-24)	0/1
Day of the week(Sunday)	0/1
Day of the week(Monday)	0/1
...	0/1
Day of the week(Saturday)	0/1
Season of the year(Summer)	0/1
...	0/1
Season of the year(Winter)	0/1

Table I
OVERALL ORACLE OUTPUT

data selector may choose the whole or the subset of the output from the Oracle. It should be noted that all the knowledge proposed by the Oracle might not be useful for a particular appliance in a house and there are different possible structuring of the knowledges proposed by the Oracle. All the outputs of the Oracle is stored in the data selector, but only those which are validated by the predictor are selected for the overall prediction. In this work, only one of the possible structuring is implemented, which is done by taking the knowledges proposed by the Oracle as a single unit after selection. There are other possible structuring which will be looked in the future.

D. Predictor

This module consists of the classifiers commonly used in Machine Learning such as the Neural networks. The classifier gets its input from the data selector. It first validates the knowledges proposed by the Oracle and then the prediction for the next 24 hours.

III. CHOICE OF CLASSIFIER AND PARAMETERS

In this section the justification of using a neural network classifier for such an application is discussed. Choice of neural networks are more on the basis of past literature than on the initial results seen in table II, where different non-linear classifiers are compared. The comparison is based on past consumption history and then prediction for the next hour. The comparison with other classifiers such as Support vector machines, Naive Bayes and K-nearest Neighbors are given. It must be mentioned, that at no point the fact that other classifiers may perform better is disregarded, these are initial results with suitable parameters. The results are the accuracy for the next hour.
In table IV the parameters of the neural network classifier is given. The number of hidden layers are chosen to be one and the number of hidden neurons to be half of the number of input nodes. This choice is to avoid the over fitting or under-fitting of the network. The choice of training algorithm is also shown in III. The results of the choice of architecture is shown in IV. The final choice of all the parameters are shown in table V.

Appliance	**SVM**	**Naive Bayes**	**KNN**	**Neural Network**
900 lamp	82.40	60.1	79.72	82.94
932 oven	84.42	84.25	83.51	84.95

Table II
CLASSIFIER COMPARISON

Training Algorithm	**Accuracy**
Gradient descent	57.20
BGFS	82.94
Conjugate entropy	83.08

Table III
TRAINING ALGORITHM

Architecture	**Accuracy**
RBF	57.20
MLP	83.22

Table IV
NEURAL NETWORK ARCHITECTURE

The scoring is done in terms of accuracy, where accuracy is the number of correct classification to the total number of classifications.

Parameter	**Selection**
Sampling Method	Random
Train sample size	75
Test sample size	25
Network Type	MLP
Activation function(hidden unit)	Tanh
Activation function(output unit)	Softmax
No of hidden neurons	no of input/2
Error Function	Cross entropy
Training Algorithm	BGFS
Learning Rate	0.1

Table V
NEURAL NETWORK PARAMETERS

IV. ORACLE KNOWLEDGE RESULTS FOR DATA SELECTION

In this section each of the proposed knowledge is validated in an incremental manner. The results indicate that all the knowledges proposed by the Oracle might not result in increase in performance of the prediction system. So for each appliance in a house a subset of the knowledge proposed by the Oracle is selected. Therefore, only knowledge which increase in prediction performance is selected for a particular appliance. All the predictions are done using a Neural Network Predictor whose parameters are discussed in III. It must be mentioned, that the knowledge proposed by the Oracle are prioritized on the basis of domain knowledge. It can be seen from table VI that due to our incremental approach the knowledge which appears first has a higher chance of getting selected than the next one.

Knowledge	**Neural Network Prediction**	**Selected**
Past consumption	82.94	✓
+ Time of the day	83.45	✓
+ Day of the week	83.73	✓
+ Season in the year	84.14	✓
+ Same hour previous 7 day	83.50	

Table VI
ORACLE RESULT : HOUSE- 900; APPLIANCE-LAMP

V. OVERALL RESULT

After the selection of the data, prediction is done for the following 24 hours at each hour and the results in terms of accuracy are shown in table VII. Here the prediction system is scored by two methods, one is by simple averaging all the accuracy for the 24 hours and the second one is a weighted average. The proposed weighing scheme is expressed by the following equation :

$$\Rightarrow \sum \frac{2(24-i)}{25*24} * Accuracy[i] \text{ for i=0,1,...,23}$$

It is a linear weighting scheme giving more importance to the first hour and least to the 24th hour. This is done due to the fact that all the hours other than the next hour will be predicted again in the next step. Results contain both the scoring methods and also the results of some trivial knowledge based predictors. The trivial knowledge based predictors are

- The predictor that always predicts the appliance wont start : Never starts.
- The predictor that always predicts the appliance will start : Always start.
- The predictor that predicts "what happens the previous day at the same hour happens the next day", i.e 24 hour similarity.
- The predictor that predicts "what happens the previous week at the same hour happens the next day", i.e 168 hour similarity.
- The predictor that predicts "what happen a random hour back happens the next hour" , i.e random hour similarity.

These estimates are important because they give an overall idea of the performance of the proposed model.

VI. DISCUSSION

The results indicates that the proposed model works better than other trivial knowledge based predictors. Previous works on appliance usage prediction from consumption data relied heavily on the assumptions expressed in the trivial knowledge based predictor. The assumption of 24 hour or 168 hour similarity is intuitive but other knowledges also need to be incorporated to make the system dynamic. The incorporation and representation of the expert knowledge helps the system to perform better as seen in the table VII. Now, from the results an overall idea about the predictability of the appliance is seen. Though it must be mentioned here that the high prediction for some appliances at homes are due to the fact that some appliances are ON or OFF at most of the time. Results show that both the categories (ON and OFF) for the appliances is predicted. Appliances which are started very few times seem to require less knowledge for prediction.

Among appliances, from the results, it indicates that the lamp requires most of the knowledges proposed by the expert among other appliances. The applicability of expert knowledge varies not only from appliance to appliance but also from House to House as the user behavior is different.

VII. CONCLUSION

To anticipate the energy needed for a service in a home automation system, the system must take into account the actions which will be done by the inhabitants. In this context, a proper prediction of energy demand in housing sector is very important. This work focuses on the prediction of the appliance usage in housing because it is a very important problem in a home automation system. The objective is to construct a model able to predict the appliance usage in housing which help the system to organize energy production and consumption and to decide which appliance will be used at each hour (energy planing). In this work we tried to predict if a particular appliance will be used at a particular hour looking 24 hours in the future. The proposed approach tried to formalize expert knowledge using predicate functions and also find a suitable data structuring for the classifier. The model is validated using an IRISE database which contains the consumption record of 100 houses for a period of 1 year. Our initial results indicate that the approach is useful in appliance usage prediction and its comparison with other trivial knowledge based predictor validates our approach. This model is applied to a wide range of appliances and houses and the initial results are encouraging.

Going into the future our aim is to build a general, fully automated and user interactive prediction system for home automation and simulate how the prediction is actually helping energy management in homes. By user interface we mean prediction also controlled by inhabitants where the users calender can be incorporated.

Appliance	Never Start	Always Start	24 hour similarity	168 hour similarity	Random hour similarity	Neural Networks (Average accuracy)	Neural Network (weighted accuracy)
900 Lamp	43.73	56.26	71.25	66.99	50.36	78.25	78.85
983 Electric Heater	71.76	28.23	94.38	89.39	92.35	96.31	96.68
925 Lamp	32.02	67.97	88.014	83.35	80.33	90.55	90.82
932 Oven	84.46	15.53	80.32	80.76	72.59	86.00	86.03
986 TV	72.13	27.86	71.36	69.43	57.30	76.64	77.18
951 Cooker	93.94	6.05	89.72	89.68	88.80	94.18	94.21
939 Washing machine	88.88	11.11	86.25	86.79	78.45	89.55	89.89

Table VII
OVERALL RESULT

REFERENCES

[1] P. Palensky and R. Posta, "Demand side management in private home using lonworks," in *Proceedings of the IEEE International Workshop on Factory Communication Systems*, 1997.

[2] K. Wacks, "The impact of home automation on power electronics," in *Applied Power Electronics Conference and Exposition*, 1993, pp. 3–9.

[3] S. Ha, H. Jung, and Y. Oh, "Method to analyze user behavior in home environment," *Personal Ubiquitous Comput.*, vol. 10, pp. 110–121, January 2006. [Online]. Available: http://dx.doi.org/10.1007/s00779-005-0016-9

[4] S. Abras, S. Ploix, S. Pesty, and M. Jacomino, "A multi-agent design for a home automation system dedicated to power management," in *Artificial Intelligence and Innovations 2007: from Theory to Applications*, ser. IFIP International Federation for Information Processing, C. Boukis, A. Pnevmatikakis, and L. Polymenakos, Eds. Springer Boston, 2007, vol. 247, pp. 233–241.

[5] S. Abras, S. Pesty, S. Ploix, and M. Jacomino, "An anticipation mechanism for power management in a smart home using multi-agent systems," in *Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on*, april 2008, pp. 1 – 6.

[6] S. Abras, S. Ploix, S. Pesty, and M. Jacomino, "Advantages of mas for the resolution of a power management problem in smart homes," in *8th International Conference on Practical Applications of Agents and Multi-Agent Systems, PAAMS'2010.* Salamanca, Spain: Springer Verlag, 26-28 April 2010.

[7] S. Abras, S. Ploix, and S. Pesty, *Housing, Housing Costs and Mortgages: Trends, Impact and Prediction*, ser. Housing Issues, Laws and Programs. Nova Publishers, 2010, no. ISBN 978-1-60741-813-9, ch. Managing Power in a Smart Home Using Multi-Agent Systems.

[8] A. M. Elmahaiawy, N. Elfishawy, and M. N. El-Dien, "Anticipation the consumed electrical power in smart home using evolutionary algorithms," in *MCIT 2010 conference*, 2010.

[9] L. D. Ha, S. Ploix, F. Wurtz, P. Perichon, and J. Merten, "Energy management system for a photovoltaic grid-connected building," in *24th EU PVSEC and 4th World Conference on Photovoltaic Energy Conversion*, Hamburg, Germany, September, 21-26 2009.

[10] L. D. Ha, S. Ploix, M. Jacomino, and H. Le Minh, *Energy Management*, ser. ISBN 978-953-307-065-0. INTECH, 2010, ch. A mixed integer programming formulation of the home energy management problem.

[11] L. Hawarah, S. Ploix, and M. Jacomino, "User behavior prediction in energy consumption in housing using bayesian networks," in *Artificial Intelligence and Soft Computing*, ser. Lecture Notes in Computer Science, L. Rutkowski, R. Scherer, R. Tadeusiewicz, L. Zadeh, and J. Zurada, Eds. Springer Berlin / Heidelberg, 2010, vol. 6113, pp. 372–379.

[12] E. A. Feinberg and D. Genethliou, "Load forecasting," in *Applied Mathematics for Restructured Electric Power Systems*, ser. Power Electronics and Power Systems, J. H. Chow, F. F. Wu, and J. Momoh, Eds. Springer US, 2005, pp. 269–285.

[13] H. Hippert, C. Pedreira, and R. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *Power Systems, IEEE Transactions on*, vol. 16, no. 1, pp. 44 –55, feb 2001.

[14] A. Bakirtzis, V. Petridis, S. Kiartzis, M. Alexiadis, and A. Maissis, "A neural network short term load forecasting model for the greek power system," *Power Systems, IEEE Transactions on*, vol. 11, no. 2, pp. 858 – 863, may 1996.

[15] D. Park, M. El-Sharkawi, I. Marks, R.J., L. Atlas, and M. Damborg, "Electric load forecasting using an artificial neural network," *Power Systems, IEEE Transactions on*, vol. 6, no. 2, pp. 442 –449, may 1991.

[16] A. Khotanzad, R. Afkhami-Rohani, and D. Maratukulam, "Annstlf-artificial neural network short-term load forecaster generation three," *Power Systems, IEEE Transactions on*, vol. 13, no. 4, pp. 1413 –1422, nov 1998.

[17] B.-J. Chen, M. wei Chang, and C.-J. Lin, "Load forecasting using support vector machines: A study on eunite competition 2001," Tech. Rep., 2001.