

# Using Conditional Random Fields to validate observations in a 4W1H paradigm

Leon Palafox, Laszlo A. Jeni, *Member, IEEE*, and Hideki Hashimoto, *Member, IEEE*

**Abstract**— Detecting human activity has been one of the main focuses in intelligent spaces. This is achieved by using a large number of sensors attached both to the humans and the environment. Yet, these systems are prone to failure due to the parallel sensing when miss firings occur. We propose a method to test and prevent the miss firings using conditional random fields, since they provide us with a tool that allows us to confirm whether the expected output or activity is likely to happen in the space or not, given the inputs of the system, which are provided by the 4W1H paradigm, that allows us to segment every piece of information in the space into 5 simple variables (Who, When, What, Where and How).

## I. INTRODUCTION

ADVANCES in networking computing, sensor technology and robotics allow us to create more convenient environments for humans. In that context the Intelligent Space (iSpace) concept was proposed. iSpace is a space that has ubiquitous distributed sensory intelligence and actuators for manipulating the space and providing useful services [1]. It can be regarded as a system that is able to support humans, i.e. users of the space, in various ways. Actuators provide services both physical and informative to humans in the space, whereas sensors are used for observing the space and gathering information [2]. In figure 1 we can see the basic layout an intelligent space usually has, where cameras, lasers and other sensors are used to provide information and robots are used to provide services.

The iSpace consists of three basic functions, "observing", "understanding", and "acting". The "Observing" function is the most important, since it will deliver the information to know what kind of services are required. Conventionally, observation has been focused only to humans and robots, but there are a large number of objects in our living environment that affect the user's behavior. In order to offer the appropriate services using the objects, not only the physical information of the object but also the object's information due to interaction with the user is needed. Such information cannot be written beforehand and is provided only by observation. In other words the relations among humans and objects are important. Thus, 4W1H, a paradigm where "when, who, what, where and how" are sensed, is used to determine this information.

L. Palafox, L. A. Jeni and H.

Hashimoto is with the Institute of Industrial Science, University of Tokyo, Japan (corresponding author to provide phone: 81+3-5452-6258; fax: 81+3-5452-6259; e-mail: leon@hlab.iis.u-tokyo.ac.jp).

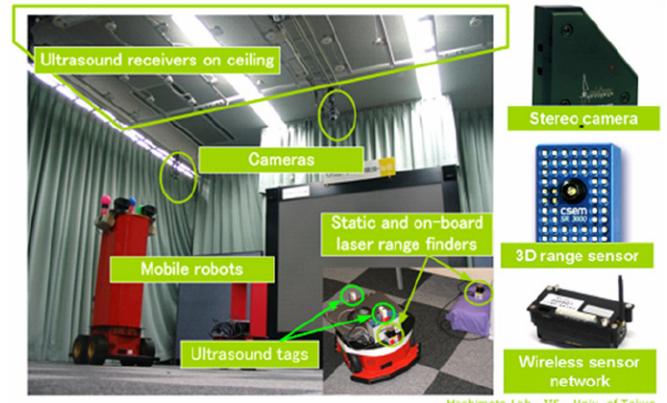


Fig. 1. Intelligent Space with Distributed Sensors and Mobile Robots.

A wide variety of sensors have been used to achieve this goal: accelerometers, positioning sensing systems, RFID tags for the users as well as the objects, etc. These sensors provide each a degree of uncertainty that has been addressed using different algorithms such as wavelets or the novel compressed sensing [3]. Yet, the results are still fallible, since users may switch activities in a fast way, making the system unable to cope with these changes and perform errors when performing the activity sensing.

There has been work done in dealing with these kind of errors, but most of it has been oriented towards addressing the problem in the sensing architecture and focuses to improve the sensitivity of the system [4]. Other approaches try to model the human activity by using Hidden Markov Models, and the overall results have been promising [5].

Based on modeling the activities as a Hidden Markov Model, we propose a new way of modeling these. Using the information obtained from the 4W1H we use Conditional Random Fields (CRF) to train the sequences [6]. CRF are a very similar model to that of Markov, with the difference that past information can be used for the further training of the system, and is a good model to fit to a human activity detection system. There has been previous work done by [7], who got good results albeit they only used cameras and a fixed set of movements.

We propose to use the 4W1H paradigm to create a sequential database that will serve us as training information for the CRF model, and afterwards we are going to apply the system in a real world environment to test its feasibility.

This paper is organized in the following way: first, we will present an overview of the 4W1H paradigm and the sensors we used to obtain the information, as well as some of the methods to obtain the parameters. Then, we will give a brief

introduction on Conditional Random Fields. We will present the experiments design as well as the testing parameters to finally show some results of the system.

## II. 4W1H

### A. Definition

Among the activities in the iSpace the one this paper is focused on is observation, and thus, an observation system that is both versatile, robust, and able to sense every significant variable that reflects at least a change in the environment is needed. Thus we need an observation system that is able in some way to relate the users with the objects, so we obtain information that is both accurate and of significance with the current activities of the human in the space. On the other hand, there is information which occurs only after a person uses an object, such as the use history that persons gives to the object, like a when a cup of coffee is used for drinking or a mouse is used for working. Such information is vast, and considering the cost it is not realistic to describe beforehand the use history information on a large number of objects that exist in the space. Therefore, it is necessary that the object's information is written automatically without human intention when the object is used by the user [8].

The paradigm itself is a data acquisition technique in which we categorize the data in our space for specific selection, in order to reduce greatly the input data and thus the processing time. It consumes a large amount of resources to track every possible variable in the environment, when the most basic variables are enough to determine user usage history and human action interpretation within a confined space. Given the last statements, we try to describe human-object relations based on following the use history of the object via the 4W1H tracking paradigm in which we declare a number of significant variables that are considered to be the most important (Figure 2) when tracking usage history of objects, those variables are:

- Where: the position of the object in a given space
- Who: the user of the object
- What: ID of the object
- When: the time of the object used
- How: the way of the object used

Each parameter provides information that allows us to know location, object interaction and human activity in the room, and since all the information is fed to a database we are able to perform an analysis of the history of use of every object, as well as the different activities human were performing at certain moments in the space.

Since the data capturing is performed online as the user generates new information, we may be able to interpret the database as a sequential database, which fits with most of the Bayesian Network Architectures, making it a feasible candidate to use architectures such as Hidden Markov models and Conditional Random Fields.



Fig. 2. 4W1H Visualization.

### B. Sensors

#### 1) How

To perform the sensing we use a MTx sensor from the company Xsens, which is a small and accurate 3DOF inertial Orientation Tracker. It provides drift-free 3D orientation as well as cinematic data: 3D acceleration, 3D rate of turn (rate gyro) and 3D earth- magnetic field . The system contains nine sensors which can be interlinked with each other in order to obtain a more complex set of data out of one specific object, as well as to provide a good architecture for setting referenced cinematic systems [9]. Using the information from the sensor we may apply classification techniques such as Self Organizing Maps to classify the movement profiles into clear activities [10].

#### 2) Where

This information is retrieved using the local IP of the computer equipped with the sensing system. Every computer in the network has a fixed physical position which is mapped in an IP list residing in the main server.

#### 3) When

The internal clock of the sensing computer will work as a master watch to keep a synchronous sensing of the system, which will be interpreted afterwards to be mapped in a more understandable space such as "Night", "Evening" or "Morning"

#### 4) Who and What

To obtain these variables, we adhere RFID tags to each of the objects and the users interacting with the system, each of these tags is equipped with an accelerometer to be able to sense in which movement a user is interacting with a new object.

## III. CONDITIONAL RANDOM FIELDS

Conditional Random Fields (CRF) are a discriminative probabilistic model that is very effective when labeling sequential data, such as natural language text or sequentially performed human activity. CRF may be considered an special case of a Hidden Markov Model (HMM) in which the

data is not represented as a Markov process, thus being past data able to provide us with information for the training phase of the algorithm.

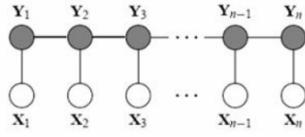


Fig. 3. Abstraction of the Conditional Random Field.

More formally, given  $X = (X_1; X_2; \dots; X_n)$  a random variable over a data sequence to be labeled  $Y = (Y_1; Y_2; \dots; Y_n)$  the joint distribution over the label  $Y$  given  $X$  has the form:

$$p_{\theta}(y|x) \propto \quad (1)$$

$$\exp \left( \sum_{e \in E, k} \vartheta_k f_k(e, y|_e, x) + \sum_{v \in V, k} \mu_k g_k(u, y|_v, x) \right) \quad (2)$$

where  $x$  is a data sequence,  $y$  a label sequence, and  $y|_s$  is a set of components of  $y$  associated with the current input  $x$ .

We assume that the features  $f_k$  and  $g_k$  are given and fixed. For this specific example, a Boolean vertex feature  $g_k$  might be true if the input  $x_i$  is performed by John and the tag  $Y_i$  is his activity.

The parameter estimation problem then becomes the estimation of the parameters  $\theta = (\vartheta_1, \vartheta_2 \dots; \mu_1, \mu_2, \dots)$  from training data with an empirical distribution (Lafferty et al., 2001).

To perform these calculations, we are using the CRF Toolbox developed for Matlab by Kevin Murphy.

#### IV. EXPERIMENTAL SETTING

To perform the experiments, we performed a series of different phases:

##### A. Data Gathering

We had a total number of 10 users performing different activities within a confined space, all of those activities were done while wearing the accelerometers in their wrists like shown in figure 2. They as well as the control objects in the environment were wearing the RFID tags to be able to identify them.

In table 1 we can observe the variety of places, times and objects that we used for the sensing, this was performed in the span of 10 days with 10 different users, it is worth to notice that not every user performed every activity or were present at every time. This was done to enhance the generalization of the system, and to be able to identify the critical variables in it, e.g Two users using a cup of coffee will use it for drinking regardless the user.

Users were changing work places at any random time, usually based on their previous activity, thus reinforcing our Conditional Random Field Model.

TABLE I  
PREDEFINED SETTING FOR THE 4W1H

User	Action	Place	Time	Object
User 1	Drink	Desk	Morning	Cup
User 2	Work	Bed	Noon	Glass
User 3	Texting	Table	Afternoon	Mouse
User 4	Read		Night	Keyboard
User 5				Mobile
User 6				Book
User 7				Magazine
User 8				
User 9				
User 10				

##### B. Training Phase

For the training part of the system, we created databases that would be feasible to happen, and inserted as well some noise, that would help the system to be trained against possible interferences within.

The training databases were drawn from a Markov Chain in which the probability of each user to change activity at any given time would be 0.3, and the probability to change places would be of 0.2.

In the training phase, we presented the CRF to a series of sequences, we performed experiments with 1, 2 and 3 sequences. We also changed the length of the sequence from 100 to 500 to analyze which were the critical variables in our system.

We performed different tests for different users, and defined the training patterns to be limited only to the What, When and How information, since most of the activity was done regardless the time or the user, and the system should be fairly capable to detect these variables without further processing in the implementation stage.

#### V. RESULTS

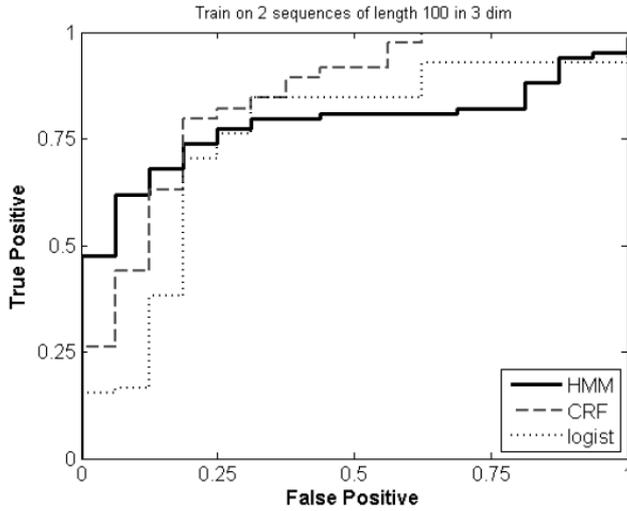
##### A. 100 Sequences

For the first set of tests, we tested the system over 100 sequences, we generated 2 and 3 sets of sequences and after the training tested the resulting chain with our experimental database. We calculated then the false positives retrieved by the algorithms as well as the correct positives.

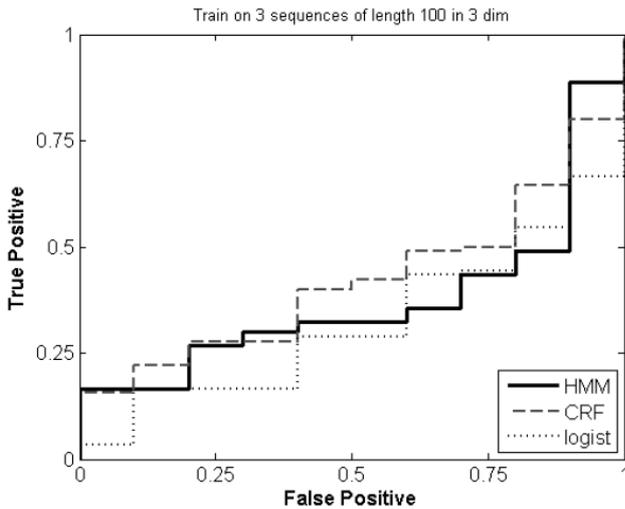
In the figure 4 we can compare the results when training 2 and 3 sets of sequences. The visualization is given in the form of a receiver operating characteristic (ROC curve) in which a high convexity towards the Y axis would mean a better performance than a straight 45° line, being the latter the middle case scenario where the systems is recognizing half of the features, in simpler words, it is guessing the result. (A 50% probability of recognition is as good as a coin toss prediction).

We as well compared our system with the results obtained using traditional Hidden Markov Models and Logistic regression for parameter classification. We can appreciate how 4a presents an almost optimal ROC curve, which means it performs a rather low number of times in a sub optimal way

(low number of false positives).



(a) ROC Curve of 2 Sequences



(b) ROC Curve of 3 Sequences

Fig. 4. Comparison of ROC curves using different amount of training sequences for 100 Patterns.

We can see how HMM and CRF perform fairly well although CRF does present an advantage overall in the test.

In 4b we see how a low number of training patterns as well as an increased number of sequences affect the system in a rather poor way, since its ROC curve is defined to be suboptimal, meaning, it is getting false triggers more than 50% of the time. We can appreciate how even in these subprime results CRF does perform a somehow better compared with the two other methods. Yet, these results present a chain that is no suitable for real time tests.

### B. 300 Sequences

In figure 5 we only present the results for a chain trained with one sequence, due to the results of the chained trained with 3 sequences resulted again in a suboptimal performance of it.

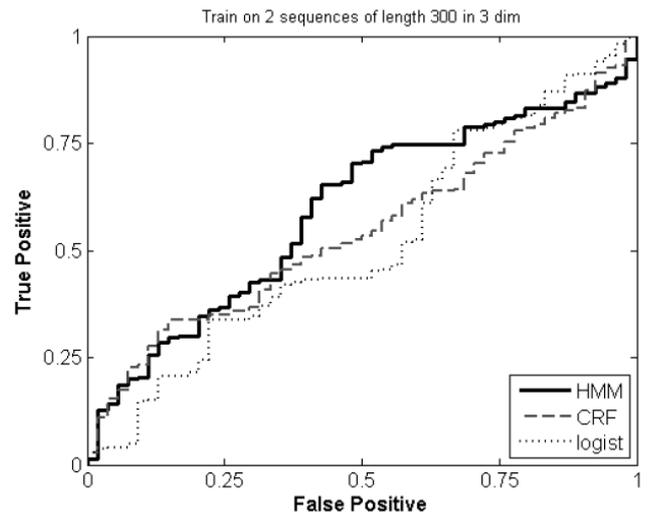
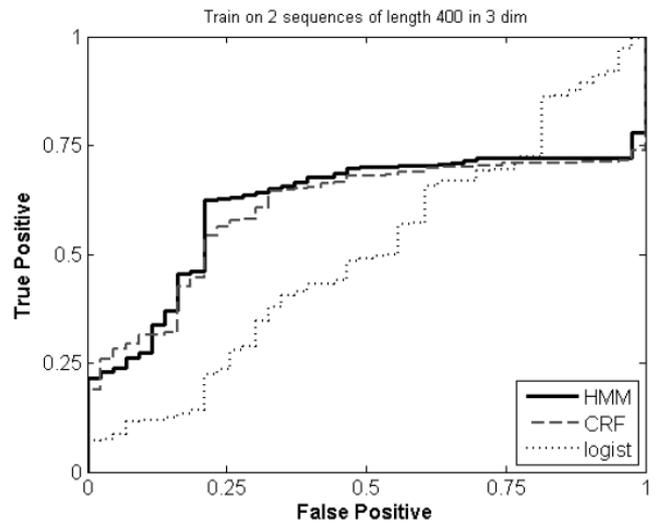
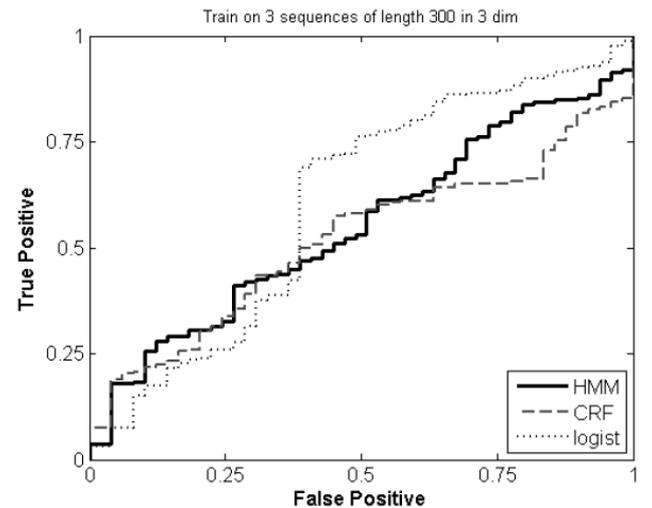


Fig. 5. ROC Curve of 2 sequences being trained for 300 Patterns.



(a) ROC Curve of 2 Sequences



(b) ROC Curve of 3 Sequences

Fig. 6. Comparison of ROC curves using different amount of training sequences for 400 Patterns.

We can appreciate how an increased number of training patterns affected the overall performance of the chain, and how algorithms like CRF and logistic regression performed in the 50% line of optimality.

While CRF also decreased its effectivity it did over performed logistic regression (being HMM the optimal tool), and the ROC curve is still a pretty good representation of the chain. The reason the system performance were affected is because a larger number of chains made the interactions among them more complex, and more unrelated activities, such as reading and sleeping may become more difficult to recognize when mixed together.

Adding and increased number of sequences, did affect the overall system in its calculation performance. Making the training times considerable longer (2x) compared with the previous example.

### C. 400 Sequences

Finally, as we can see in 6 CRF and HMM performed fairly similar, both with irregular ROC curves for the case of 2 training sequences, while logistic regression was well over the guessing line. For the training with 3 sequences rather than with 2, we do see an improvement presented by CRF that HMM and LR could not perform. These means that having 400 sequences works better for our system than having 300, yet having 100 was the best optimal performance we could obtain in repeated tests. Using Conditional Random Fields to validate observations in a 4W1H paradigm

Our system presented good results for a small number of sequences and for a small number of training patterns, as well it presented good results when dealing with the untrained patterns that were the database taken on site.

## VI. CONCLUSIONS

We have presented an implementation of Failure detection in a Human Activity detection system, in which we trained a Conditional Random Field to be able to cope with unlikely sequences, which in a CRF would presents as sequences with low probability. We have implemented these by using the 4W1H paradigm to generate the necessary features the chain needs to perform its training. We have presented that an optimal ROC curve could be obtained by training only 100 sequences, which present a feasible number for a human activity detection system, since we cannot rely on buffered data for making assumptions on the current state of the individual. In the future we still need to increase the detection ratio by means of an optimization of the CRF parameters, as well, we may include other variables such as a reinforcement observer that will allow us to do an online training along with the testing of the system.

## REFERENCES

- [1] J. H Lee and H. Hashimoto. Intelligent space concept and contents. *Advanced Robotics*, 16 (3):265{280, 2002.
- [2] D. Brscic and H. Hashimoto. Tracking of humans inside intelligent space using static and mobile sensors. In *Proceedings of the 33th*

Annual Conference of the IEEE Industrial Electronics Society (IECON'07), pages 10{15, 2007.

- [3] L. Palafox and H. Hashimoto. A compressive sensing approach to the 4w1h architecture. In *Industrial Technology (ICIT), 2010 IEEE International Conference on*, pages 1599
- [4] Seon-Woo Lee and K. Mase. Activity and location recognition using wearable sensors. *Pervasive Computing, IEEE*, 1(3):24 { 32, 2002. ISSN 1536-1268. doi: 10.1109/MPRV.2002.1037719.
- [5] M. Philipose, K.P. Fishkin, M. Perkowitz, D.J. Patterson, D. Fox, H. Kautz, and D. Hahnel. Inferring activities from interactions with objects. *Pervasive Computing, IEEE*, 3(4):50 { 57, oct.-dec. 2004. ISSN 1536-1268. doi: 10.1109/MPRV.2004.7.
- [6] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Machine Learning International Workshop*, pages 282 { 289, 2001.
- [7] C. Sminchisescu, A. Kanaujia, Zhiguo Li, and D. Metaxas. Conditional models for contextual human motion recognition. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1808 {1815 Vol. 2, 17-21 2005. doi: 10.1109/ICCV.2005.59.
- [8] Mihoko Niitsuma, Kazuki Yokoi, and Hideki Hashimoto. Describing human-object interaction in intelligent space. In *HST'09: Proceedings of the 2nd conference on Human System Interactions*, pages 392{396, Piscataway, NJ, USA, 2009. IEEE Press. ISBN 978-1-4244-3959-1.
- [9] Xsens. Specification sheet. World Wide Web electronic publication, 2006. URL <http://www.xsens.com>.
- [10] L. Palafox and H. Hashimoto. A Movement Profile Detection System Using Self Organized Maps in the Intelligent Space. In *Tokyo, Japan: IEEE Workshop on Advanced Robotics and its Social Impacts*, page 114, 2009.