HomeActivity: Recognizing home activities using sensor data

Stephen Lee, Patrick Pagus and Dong Chen

December 16, 2015

1 Abstract:

Home activities recognizing allows many potencial smart applications including healthcare and energy efficiency. In this project, we present the state of the art models for recognizing activities, and show how they performance on the Kasteren-a real world dataset. Comparative analysis metrics including precision, recall, accuracy, F1 mesure and MCC are used in the experiments. Our evaluation results could be used for other pattern recognizing area rather than home activity recognizing.

2 Introduction

Automatically recognizing human activities such as cooking, sleeping and bathing, in a home setting allows many applications in areas such as intelligent environments and healthcare. In healthcare, home activity recognition can be used to assess the cognitive and physical capabilities of an elderly person. Home energy efficiency can be improved given a schedule of the occupant's activities by scheduling background loads, such as heating, cooling, and various appliances.

Recent developments in sensing technology has led to wireless sensor networks which provide a non-intrusive, privacy friendly and easy to install solution to in-home monitoring. Sensors used are contact switches to measure open-close states of doors and cupboards; pressure mats to measure sitting on a couch or lying in bed; mercury contacts for movement of objects such as drawers; passive infrared sensors to detect motion in a specific area and float sensors to measure the toilet being flushed. More formally, activity recognition finds a function that maps a sequence of sensor readings (feature vectors), to a sequence of activity labels that best describes the sensor data. We focus on activity recognition within home based wireless sensor networks. This domain has several challenges. First, there is a dearth of training data because labeling one's activities is a tedious time consuming process, this limits the number of effective techniques. Second, patterns in training data are specific to the home and the individual. Therefore, the performance of the different

machine learning technique may vary widely over a real world dataset. Finally, dominant classes may confound machine learning techniques.

To further our knowledge of machine learning, we will conduct a comprehensive empirical analysis of proposed methods in this domain. These methods include Hidden Markov Models(HMM), Conditional Random Fields(CRF), Support Vector Machines(SVM), and Structured Support Vector Machines(SSVM) and Naive Bayes (NB). We implemented the models in Python and apply these methods to Kasteren-a real world dataset. By applying these techniques to real-world datasets and explaining the differences and similarities in their performances, we will gain a deeper understanding of machine learning. Furthermore, Our evaluation results could be used as a baseline for other pattern recognizing area rather than home activity recognizing.

3   Related Works

The probabilistic models discussed in our project represent the state of the art models used in activity recognition. Tapia et al. used the naive Bayes model in combination with the raw feature representation on two real world datasets recorded using a wireless sensor network [1]. HMMs were used in work by Patterson et al. and were applied to data obtained from a wearable RFID reader in a house where many objects are equipped with RFID tags [2]. In work by van Kasteren et al. the performance of HMMs and CRFs in activity recognition was compared on a realworld dataset recorded using a wireless sensor network [3]. Duong et al. compared the performance of HSMMs and HMMs in activity recognition using a laboratory setup in which four cameras captured the location of a person in a kitchen setup [4]. One type of model that we have not included in our comparison are hierarchical models. They have been successfully applied to activity recognition from video data [5], in an office environment using cameras, microphones and keyboard input [6] and on data obtained from a wearable sensing system [7]. The related works show part of the models in their work, respectively. It is hard to compare across different works. Instead, we implemented a set of models and ran the experiments on the same real word dataset. The evaluation results could be used as the baseline for future research.

4   Research Problem

We use more models, compared with other related works.

5   Recognizing Methods

We would like to explore the following techniques:

1. HMM - Hidden Markov Modes (HMM) are a ubiquitous tool for modeling time series data and a generative probabilistic model consisting of a hidden variable and an observable variable at each time step. HMM can be used as black-box density models on sequences. They have the advantage over Markov models in that they can represent long-range dependencies between observations, mediated via the latent variables. In our project the hidden variable is the activity performed, and the observable variable is the vector of sensor readings. There are two dependency assumptions: The hidden variable at time t, namely y_t, depends only on the previous hidden variable y_{t-1}. The observable variable at time t, namely x_t, depends only on the hidden variable y_t at that time slice. Then, we can specify an HMM using three probability distributions: the distribution over initial states p(y_1); the transition distribution $p(y_t \mid y_{t-1})$ representing the probability of going from one state to the next; and the observation distribution $p(x_t \mid y_t)$ indicating the probability that the state yt would generate observation x_t. Learning the parameters of these distributions corresponds to maximizing the joint probability p(x, y) of the paired observation and label sequences in the training data. We can factorize the joint distribution in terms of the three distributions described above as follows:

$$P(x, y) = \prod_{t=1}^{T} p(y_t \mid y_{t-1}) p(x_t \mid y_t), p(y_1) = p(y_1 \mid y_0) \quad (1)$$

The parameters that maximize this joint probability are found by frequency counting. Because in our case we are dealing with discrete data, we can simply count the number of occurrences of transitions, observations and states [1].

2. CRF - A CRF using the first-order Markov assumption is called a linear-chain CRF and most closely resembles the HMM in terms of structure. We define the linear-chain CRF as a discriminative analog of the previously defined HMM, so that the conditional distribution is defined as

$$p(y \mid x) = \frac{1}{Z(x)} exp \left[\sum_{k=1}^{K} \lambda_k f_k(y_t, y_{t-1}, x_t) \right] \quad (2)$$

where K is the number of feature functions used to parameterize the distribution, λ_k is the weight parameter and $f_k(y_t, y_{t-1}, x_t)$ is the feather function. The product of the parameters and the feature function $\lambda_k f_k(y_t, y_{t-1}, x_t)$ is called energy function, while the exponential representation is the potential function. The partiton function $Z(x)$ is normalization term in order to make the distribution sums up to one and obtains a probabilistic interpretation.

$$Z(x) = \sum_{y} exp \left[\sum_{k=1}^{K} \lambda_k f_k(y_t, y_{t-1}, x_t) \right] \quad (3)$$

3

3. SVM - The implementation used in this experiment comes from SVM

4. SSVM -

5. Naive Bayes. The naive Bayes model is one of the most simplistic probabilistic models. Unlike the other models, the naive Bayes model assumes all data points are independently and identically distributed (i.i.d.), that is it does not take into account any temporal relations between datapoints. The model factorizes the joint probability over the datapoints as follows,

$$p(y_{1:T}, x_{1:T}) = \prod_{t=1}^{T} p(x_t \mid y_t)p(y_t) \tag{4}$$

we apply the naive Bayes assumption, which means we model each sensor reading separately, requiring only N parameters for each activity. The observation distribution therefore factorizes as

$$p(x_t \mid y_t = i) = \prod_{n=1}^{N} p(x_t^n \mid y_t = i) \tag{5}$$

where each sensor observation is modeled as an independent Bernoulli distribution, given by

$$p(x_t^n \mid y_t = i) = \mu_{ni}{}^{x_t{}^n}(1 - \mu_{ni})^{1-x_t{}^n} \tag{6}$$

6 Experimental Evaluation

6.1 Datasets

The Kasteren dataset is recording a 26-year-old man. He lives alone in a three-room apartment where 14 state-change sensors were installed. Locations of sensors include doors, cup- boards, refrigerator and a toilet flush sensor. Sensors were left unattended, collecting data for 28 days in the apartment. This resulted in 2120 sensor events and 245 activity instances.

6.2 Experimental Setup

6.3 Feather Representation

Raw
Changepoint
Last-fired

6.4 Comparative Analysis Metrics

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$F - Measure = \frac{2 \cdot Precision \cdot Recall}{Precison + Recall} \tag{10}$$

The Matthews correlation coefficient is used in machine learning as a measure of the quality of binary (two-class) classifications, introduced by biochemist Brian W. Matthews in 1975. It takes into account true and false positives and negatives and is generally regarded as a balanced measure which can be used even if the classes are of very different sizes. The MCC is in essence a correlation coefficient between the observed and predicted binary classifications; it returns a value between 1 and +1. A coefficient of +1 represents a perfect prediction, 0 no better than random prediction and 1 indicates total disagreement between prediction and observation.

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \tag{11}$$

6.5 Experiments

7 Discussion and Future Work

8 Reference

References

[1] E. M. Tapia, S. S. Intille, and K. Larson. *Activity recognition in the home using simple and ubiquitous sensors.* In Pervasive Computing, Second International Conference, PERVASIVE 2004, pp. 158175, Vienna, Austria (April, 2004).

[2] D. J. Patterson, D. Fox, H. A. Kautz, and M. Philipose. Fine-grained activity recognition by aggregating abstract object usage. In ISWC, pp. 4451. IEEE Computer Society, (2005). ISBN 0-7695-2419-2. URL http://doi.ieeecomputersociety.org/10.1109/ISWC.2005.22.
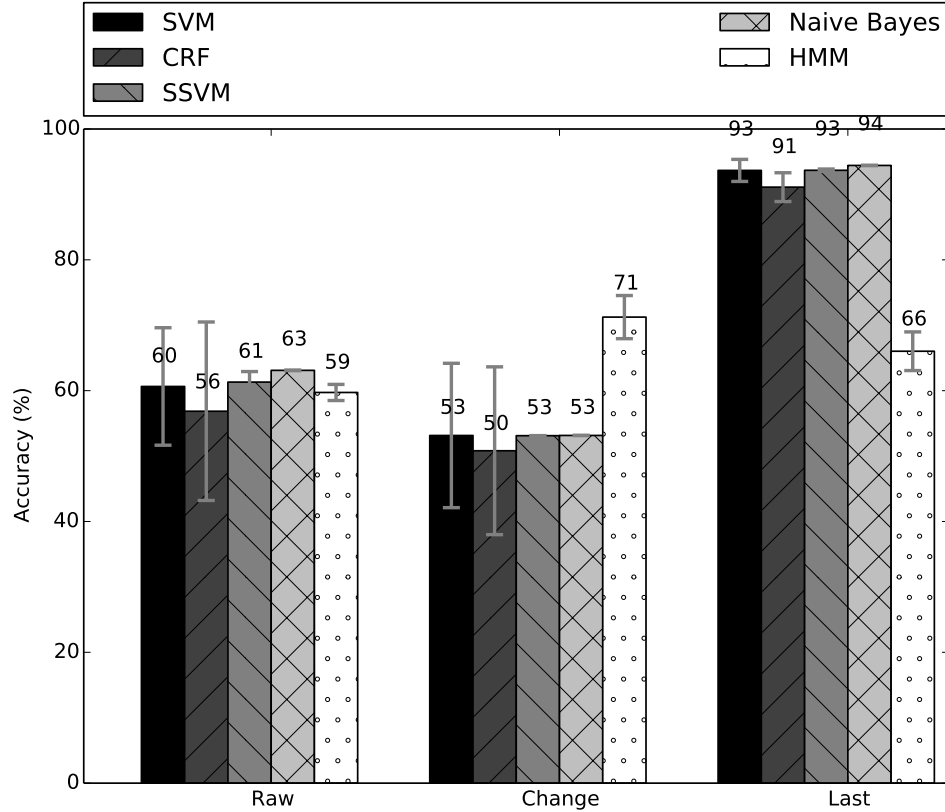
Figure 1: House A

[3] T. van Kasteren, A. Noulas, G. Englebienne, and B. Kro se. Accurate activity recognition in a home setting. In UbiComp 08: Proceedings of the 10th international conference on Ubiquitous computing, pp. 19, New York, NY, USA, (2008). ACM. ISBN 978-1-60558-136-1. doi: http: //doi.acm.org/10.1145/1409635.1409637.

[4] T. Duong, D. Phung, H. Bui, and S. Venkatesh, Efficient duration and hierarchical modeling for human activity recognition, Artif. Intell. 173(7-8), 830856, (2009). ISSN 0004-3702. doi: http://dx.doi.org/10.1016/j.artint.2008.12.005.

[5] S. Luhr, H. H. Bui, S. Venkatesh, and G. A. West, Recognition of human activity through hierarchical stochastic learning, percom. 00, 416, (2003). doi: http://doi.ieeecomputersociety. org/10.1109/PERCOM.2003.1192766.

[6] N. Oliver, A. Garg, and E. Horvitz, Layered representations for learning and inferring office activity from multiple sensory channels, Comput. Vis. Image Underst. 96(2), 163180, (2004). ISSN 1077-3142. doi: http://dx.doi.org/10.1016/j.cviu.2004.02.004.
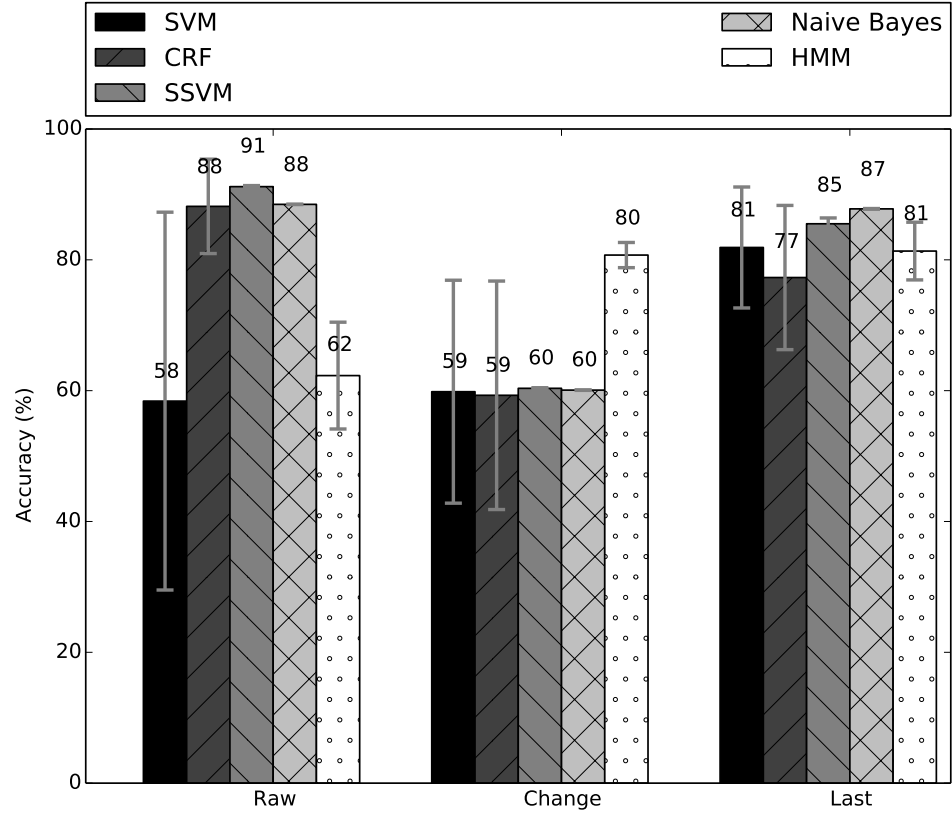
Figure 2: House B

[7] A. Subramanya, A. Raj, J. Bilmes, and D. Fox. Hierarchical models for activity recognition. In IEEE Multimedia Signal Processing (MMSP) Conference, Victoria, CA (October, 2006).
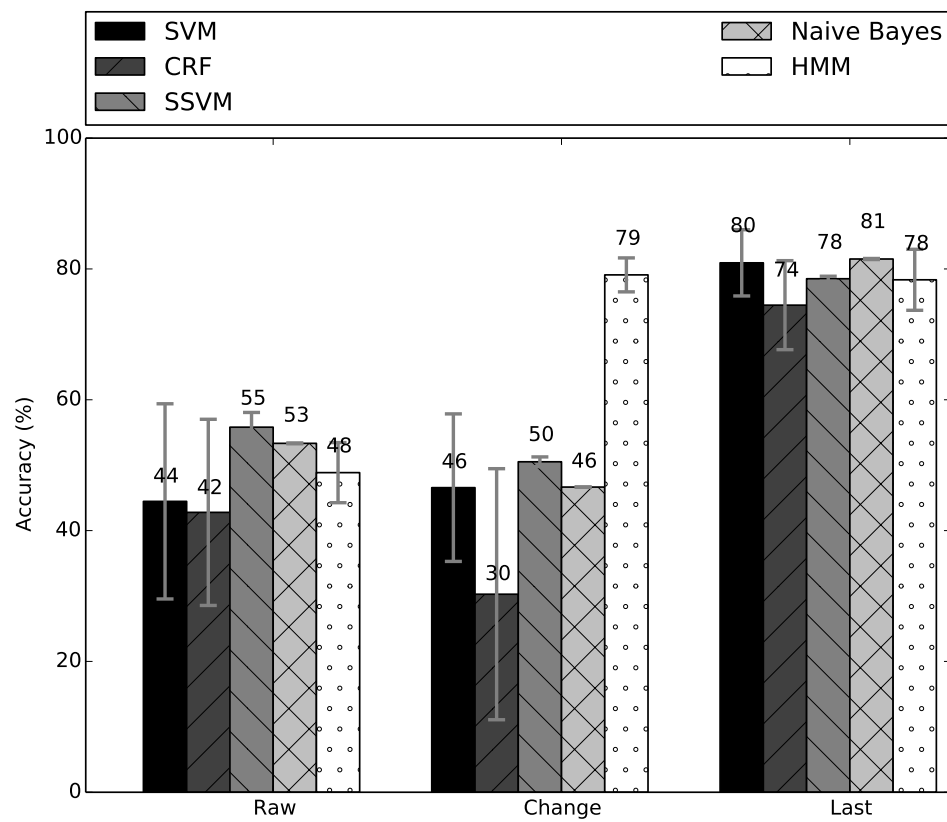
Figure 3: House C