Smart Walkers! Enhancing the Mobility of the Elderly

(Extended Abstract)

Mathieu Sinn
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
msinn@cs.uwaterloo.ca

Pascal Poupart
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
ppoupart@cs.uwaterloo.ca

ABSTRACT

The idea of Smart Walkers is to equip customary rolling walkers with sensors in order to assist users, caregivers and clinicians. The integral part of the Smart Walkers is an autonomous agent which monitors the activity of the user, assesses his physical conditions, and detects potential risks of falls. In this paper, we study methods which enable the agent to recognize the user activity from the sensor measurements. The proposed methods use Conditional Random Fields with features based on discriminant rules. A special case are features which, in order to distinguish between two activities, compare the sensor measurements to thresholds learned by a linear classifier. Experiments with real user data show that the methods achieve a good accuracy; the best results are obtained using "smooth" thresholds based on sigmoid functions.

Categories and Subject Descriptors

J.3 [**Computer Applications**]: Life and Medical Sciences—*Health*; I.2.6 [**Computing Methodologies**]: Robotics—*Sensors*

General Terms

Experimentation

Keywords

Single agent learning, Reasoning

1. INTRODUCTION

Safe and independent mobility is a key factor in the quality of life of elderly people. Mobility aids, such as canes, rolling walkers and wheel chairs, encourage independent mobility, however, improper use can induce additional risks of falling, particularly as the individual motoric capabilities deteriorate. To improve the utility of mobility aids, we are developing a mixed-initiative system, called Smart Walker, which is a customary four-wheel rolling walker equipped with a set of sensors. The integral part of the Smart Walker is an autonomous agent which takes into account the sensor measurements and monitors the user activity. Our goal is to assist users, caregivers and clinicians, e.g., by monitoring the user's stability, supervising the execution of daily excercises and providing longitudinal data of the physical and mental conditions of walker users. A key step in implementing these functionalities is enabling the agent to recognize the activity of the user from the sensor measurements.

Cite as: Smart Walkers! Enhancing the Mobility of the Elderly (Extended Abstract), Mathieu Sinn and Pascal Poupart, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems – Innovative Applications Track (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. 1133-1134.

2. ACTIVITY RECOGNITION

We use the following sensor measurements: x_t^{speed}, the speed of the walker; $x_t^{\text{tot. load}}$, the total load on the four wheels; x_t^{FCOP}, the relative difference between the load on the left and the right wheels; x_t^{SCOP}, the difference between the load on the rear and the front wheels; $x_t^{\text{x-acc.}}$, $x_t^{\text{y-acc.}}$ and $x_t^{\text{z-acc.}}$, the acceleration in the three spatial dimensions. In order to include information on the past, we also compute the mean and the variance over the previous 5 and 25 time points. Note that the measurements are digitized with 50 Hz, so 25 time points correspond to half a second.

2.1 Conditional Random Fields

In [3], we compared the performance of several probabilistic models and found that the best results were obtained for Conditional Random Fields (CRFs). A CRF specifies the distribution of a sequence of labels, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, conditional on a sequence of observations, $\boldsymbol{X} = (X_1, \ldots, X_n)$ (see [2]). In our context, the observations represent the sensor measurements, and the hidden states the user activities. CRFs are parameterized by features, $\boldsymbol{f}$, and model weights, $\boldsymbol{\lambda}$. For any $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$, the probability of $\boldsymbol{Y} = \boldsymbol{y}$ conditional on $\boldsymbol{X} = \boldsymbol{x}$ is given by

$$P_{\boldsymbol{\lambda}}(\boldsymbol{Y} = \boldsymbol{y} \,|\, \boldsymbol{X} = \boldsymbol{x}) \quad \propto \quad \exp\big(\boldsymbol{\lambda}^T \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})\big).$$

For the labeling of sequential data, linear-chain CRFs are of particular importance. For that type of models, $\boldsymbol{\lambda}^T \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})$ can be written in terms of state and transition features:

$$\boldsymbol{\lambda}^T \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) \quad = \quad \sum_{t=1}^{n} \boldsymbol{\mu}^T \boldsymbol{f}^{\text{state}}(x_t, y_t) + \sum_{t=2}^{n} \boldsymbol{\nu}^T \boldsymbol{f}^{\text{trans}}(y_{t-1}, y_t).$$

More generally, $\boldsymbol{f}^{\text{trans}}$ may also depend on x_t. In our experiments, we chose $\boldsymbol{\nu}^T \boldsymbol{f}^{\text{trans}}(y_{t-1}, y_t) = \nu\, \mathbf{1}(y_{t-1} = y_t)$, which simply reflects whether or not an activity persists. For the selection of the state features, we propose to use *discriminant rules*. The basic idea is, in order to determine the compatibility of the events $X_t = x_t$ and $Y_t = i$, to consider any potential alternative, $Y_t = j$, and to assess whether

$X_t = x_t$ is more compatible with $Y_t = i$ or $Y_t = j$. Writing $\mathcal{Y}$ for the set of all labels, this gives us

$$\boldsymbol{\mu}^T \boldsymbol{f}^{\text{state}}(x_t, i) \quad = \sum_{j \in \mathcal{Y} \setminus \{i\}} \boldsymbol{\mu}_{ij}^T \boldsymbol{d}_{ij}(x_t),$$

where $\boldsymbol{d}_{ij}(\cdot)$ are functions discriminating between i and j, associated with the weights $\boldsymbol{\mu}_{ij}$. In the following, we consider several examples.

2.1.1 Binary Thresholds

The simplest type of discriminant rules is obtained by comparing the observations (component-wise) to thresholds. Write $\mathbf{1}(\cdot)$ for the function evaluating to 1 if the statement in the brackets is true and to 0, otherwise. Then

$$\boldsymbol{\mu}_{ij}^T \boldsymbol{d}_{ij}(x_t) \quad = \quad \mu_{ij}^{(g)} \mathbf{1}(x_t \geq \tau_{ij}) + \mu_{ij}^{(l)} \mathbf{1}(x_t < \tau_{ij}).$$

For the selection of τ_{ij}, suppose that we are given training data $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$. Write n_i for the number of points in the training data for which $y_t = i$, and let $\mu_i := \frac{1}{n_i} \sum_{t=1}^{n} \mathbf{1}(y_t = i) x_t$. Similarly, define n_j and μ_j. In our experiments, we use the threshold $\tau_{ij} = (\mu_i + \mu_j)/2$. Note that τ_{ij} is the threshold obtained by Linear Discriminant Analysis if n_i and n_j are equal (see [1]).

2.1.2 Sigmoid Thresholds

In order to take into account by what margin x_t exceeds τ_{ij}, we consider continuous thresholds based on the sigmoid function $\text{sig}(x) = 1/(1 + e^{-x})$. The slope is determined by a scaling parameter γ_{ij}, yielding

$$\boldsymbol{\mu}_{ij}^T \boldsymbol{d}_{ij}(x_t) \;=\; \mu_{ij}^{(g)} \text{sig}(\gamma_{ij}(x_t - \tau_{ij})) + \mu_{ij}^{(l)} \text{sig}(\gamma_{ij}(\tau_{ij} - x_t)).$$

Note that the larger γ_{ij}, the more similar are the continuous thresholds to the binary ones. For the selection of γ_{ij}, we maximize the likelihood of a logistic regression model with the slope γ_{ij} and the intercept $-\gamma_{ij}\tau_{ij}$ (see [1]).

2.1.3 Using Raw Observations

Finally, we consider discriminant rules based on the raw observations. Let μ and σ^2 denote the sample mean and variance of x_t in the training set. Then we use the rules

$$\boldsymbol{\mu}_{ij}^T \boldsymbol{d}_{ij}(x_t) \quad = \quad \mu_{ij}^{(ic)} + \mu_{ij}^{(sl)} \big(\sigma^{-1}(x_t - \mu) \big).$$

The standardization of x_t is necessary to avoid a penalization of $\mu_{ij}^{(ic)}$ and $\mu_{ij}^{(sl)}$ during the training of the CRF.

3. EXPERIMENTS

We collected user data in two different setups. In the first experiment, we asked 12 healthy young subjects (19-53 years old) to walk twice through a predefined course which included the following activities: not touching the walker (**N**), stopping (**S**), walking forward/backwards (**F**/**B**), turning left/right (**L**/**R**), transferring between the walker and a chair (**T**). The participants of the second experiment were 15 older adults (80-97 years old), 8 of which were regular walker users. Besides the activities in the first experiment, the participants were sitting on the walker (**SI**), going up/down a ramp (**UR**/**DR**), and going up/down a curb (**UC**/**DC**). While they were performing the courses, we asked the participants of the second experiment to execute real-life tasks like picking up objects from the ground or walking at different speeds; moreover, we recorded some spontaneous activity in between the two courses.

Table 1: Accuracy for Experiment 1 (in %)

	N	S	F	L	R	B	T	Tot.
Thre	81	70	95	74	65	91	61	87
Sigm	88	71	96	77	71	92	56	89
Raw	91	57	96	71	60	88	42	86
Bin	75	73	95	74	67	92	53	86

Table 2: Accuracy for Experiment 2 (in %)

	S	F	L	R	SI	UR	DR	UC	DC	Tot.
Thre	89	82	56	52	98	65	54	60	55	81
Sigm	90	85	63	51	99	79	58	61	54	83
Raw	89	85	58	46	99	67	63	55	47	82
Bin	89	85	58	53	99	72	52	56	58	82

We compare four different methods: **Thre**, based on binary thresholds; **Sigm**, based on sigmoid thresholds; **Raw**, using the raw observations; **Bin**, using features based on data binning, where we chose the number of data bins equal to the number of different labels. Given the trained CRF and observations $\boldsymbol{x} = (x_1, \ldots, x_n)$, we predict the sequence of labels $\boldsymbol{y} = (y_1, \ldots, y_n)$ component-wise by maximizing the marginal distribution of Y_t conditional on $\boldsymbol{X} = \boldsymbol{x}$.

The results are shown in Table 1 and 2. As can be seen, Sigm achieves the best performance with an overall accuracy of 89% and 83%. Except for Bin in Experiment 2, the differences in the performance are all statistically significant (one-tailed Wilcoxon test, $\alpha = 0.05$). Not surprisingly, all methods have problems to recognize transferring, which is an intermediate activity between not touching the walker and stopping. Turns are sometimes confused with walking forward, however, also for human observers it is not easy to tell when a turn exactly starts or ends.

Overall, the results for Experiment 1 are better than for Experiment 2. One reason is that the participants in Experiment 1 performed the course twice, so the training set always includes one recording of the person for which the activity is predicted. Furthermore, the activities in Experiment 2 are more individual, e.g., the participants used very different strategies to go up and down the curb. Even for simple activities the variability in Experiment 2 is higher, as the participants were instructed to perform different real-life tasks meanwhile.

4. ACKNOWLEDGMENTS

Mathieu Sinn was supported by a Postdoctoral Research Fellowship of the Canadian Government.

5. REFERENCES

[1] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning. Data Mining, Inference and Prediction.* Springer, New York, NY, 2nd edition, 2009.

[2] J. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the IEEE International Conference on Machine Learning*, 2001.

[3] F. Omar, M. Sinn, J. Truszkowski, P. Poupart, J. Tung, and A. Caine. Comparative analysis of probabilistic models for activity recognition with an instrumented walker. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*, 2010.