Activity Recognition in Smart Homes

Tom Diethe
tdiethe@amazon.com

University of Bristol, December 2017

Outline

The SPHERE project

Environmental

- temp, light, humidity, air quality, water & electricity

Video

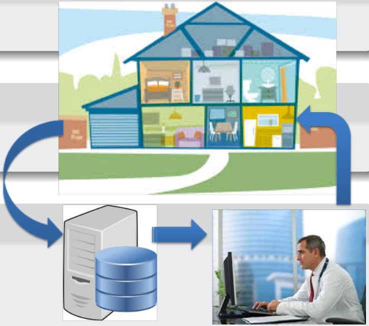
- emotion, gait, activity, interaction

Wearable

- activity, sleep, etc.

Contextual information

- demographics, medical history
- diaries, annotations



The SPHERE project

Environmental

- temp, light, humidity, air quality, water & electricity

Video

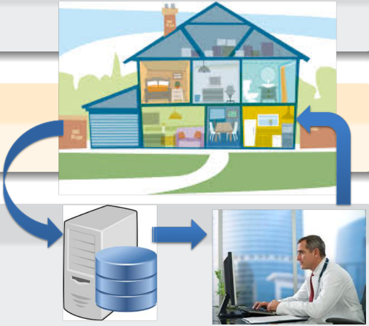
- emotion, gait, activity, interaction

Wearable

- **activity, sleep, etc.**

Contextual information

- demographics, medical history
- diaries, annotations



Goal: Activity Recognition in Smart Homes

- Noisy data
  $\longrightarrow$ methods that characterise the noise

Goal: Activity Recognition in Smart Homes

- Noisy data
  $\longrightarrow$ methods that characterise the noise
- Missing packets
  $\longrightarrow$ robust methods

Goal: Activity Recognition in Smart Homes

- Noisy data
  $\longrightarrow$ methods that characterise the noise
- Missing packets
  $\longrightarrow$ robust methods
- Limited storage and transmission capacity
  $\longrightarrow$ sparse/compressive methods

Goal: Activity Recognition in Smart Homes

- Noisy data
  $\longrightarrow$ methods that characterise the noise
- Missing packets
  $\longrightarrow$ robust methods
- Limited storage and transmission capacity
  $\longrightarrow$ sparse/compressive methods
- Labelled data time-consuming and costly to acquire

Goal: Activity Recognition in Smart Homes

- Noisy data
  $\longrightarrow$ methods that characterise the noise
- Missing packets
  $\longrightarrow$ robust methods
- Limited storage and transmission capacity
  $\longrightarrow$ sparse/compressive methods
- Labelled data time-consuming and costly to acquire
  $\longrightarrow$ Active Learning

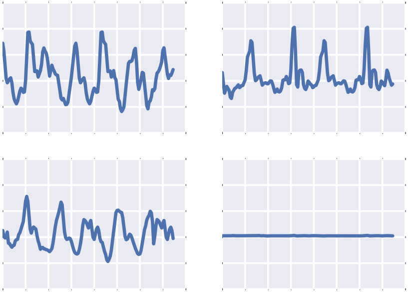
Goal: Activity Recognition in Smart Homes

- Noisy data
  $\longrightarrow$ methods that characterise the noise
- Missing packets
  $\longrightarrow$ robust methods
- Limited storage and transmission capacity
  $\longrightarrow$ sparse/compressive methods
- Labelled data time-consuming and costly to acquire
  $\longrightarrow$ Active Learning
- Differing training and deployment settings

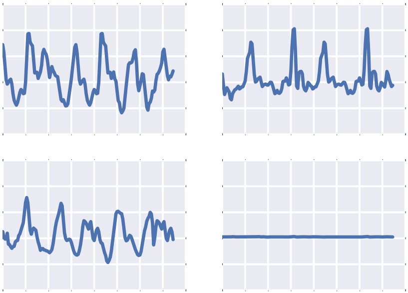
Goal: Activity Recognition in Smart Homes

- Noisy data
  $\longrightarrow$ methods that characterise the noise
- Missing packets
  $\longrightarrow$ robust methods
- Limited storage and transmission capacity
  $\longrightarrow$ sparse/compressive methods
- Labelled data time-consuming and costly to acquire
  $\longrightarrow$ Active Learning
- Differing training and deployment settings
  $\longrightarrow$ Transfer Learning

Motivation

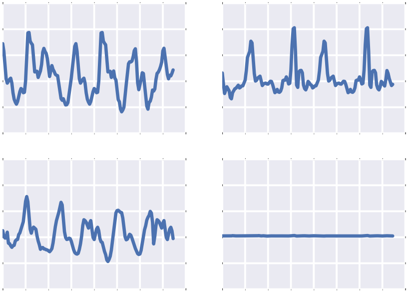


Motivation



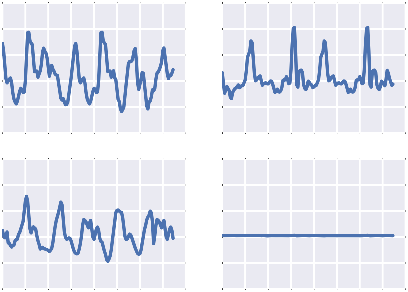
- Would like a sparse representation of the data

Motivation



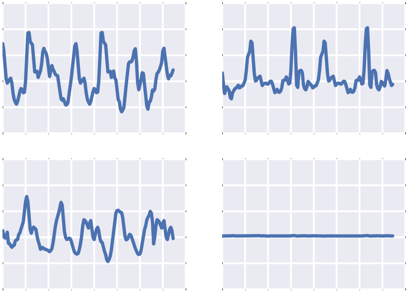
- Would like a sparse representation of the data
- Don't want to make parametric assumptions

Motivation



- Would like a sparse representation of the data
- Don't want to make parametric assumptions
- Features for classification (activity recognition)

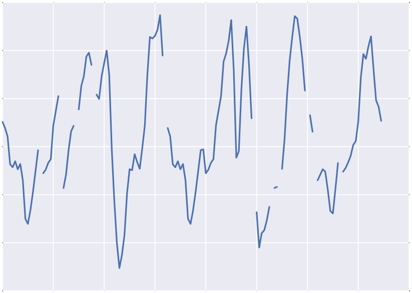
Motivation



- Would like a sparse representation of the data
- Don't want to make parametric assumptions
- Features for classification (activity recognition)

$\longrightarrow$ Dictionary Learning

Motivation

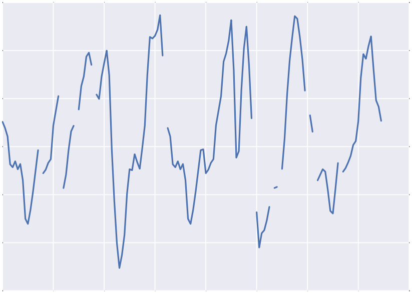


- Would like a sparse representation of the data
- Don't want to make parametric assumptions
- Features for classification (activity recognition)

  $\longrightarrow$ Dictionary Learning

- Would like to incorporate with other models

Motivation



- Would like a sparse representation of the data
- Don't want to make parametric assumptions
- Features for classification (activity recognition)

  $\longrightarrow$ Dictionary Learning

- Would like to incorporate with other models

  $\longrightarrow$ Bayesian approach

Dictionary Learning

- **Aim:** find a set of vectors $\mathbf{d}_i$, (dictionary), to represent $\mathbf{x} \in \mathbb{R}^n$ as a linear combination of these vectors:

$$\mathbf{x} = \sum_{i=1}^{k} \mathbf{z}_i \mathbf{d}_i \qquad \text{s.t.} \quad k \gg n. \tag{1}$$

Dictionary Learning

- **Aim:** find a set of vectors $\mathbf{d}_i$, (dictionary), to represent $\mathbf{x} \in \mathbb{R}^n$ as a linear combination of these vectors:

$$\mathbf{x} = \sum_{i=1}^{k} \mathbf{z}_i \mathbf{d}_i \qquad \text{s.t.} \quad k \gg n. \tag{1}$$

- **Sparsity:** few non-zero components $\mathbf{z}_i$ (or many close to 0)

Dictionary Learning

- **Aim:** find a set of vectors $\mathbf{d}_i$, (dictionary), to represent $\mathbf{x} \in \mathbb{R}^n$ as a linear combination of these vectors:

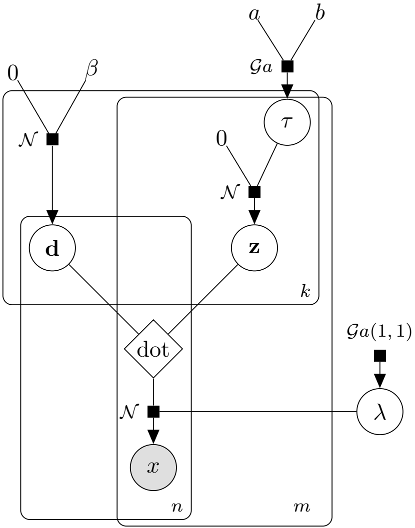
$$\mathbf{x} = \sum_{i=1}^{k} \mathbf{z}_i \mathbf{d}_i \qquad \text{s.t.} \quad k \gg n. \tag{1}$$

- **Sparsity:** few non-zero components $\mathbf{z}_i$ (or many close to 0)
- Cost function for m input vectors arranged in columns of matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$

$$\min_{\mathbf{Z},\mathbf{D}} \|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 + \lambda \sum_{i=1}^{n} \Omega(\mathbf{z}_i)$$
$$\text{s.t.} \, \|\mathbf{d}_i\|^2 \leq C, \quad \forall i = 1, \ldots, k.$$

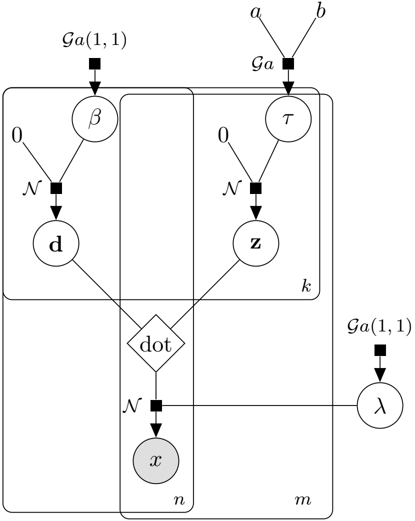
where $\mathbf{D} \in \mathbb{R}^{n \times k}$ is the dictionary, $\mathbf{Z} \in \mathbb{R}^{k \times m}$ are the coefficients, and $\Omega(\cdot)$ is a sparsity inducing regulariser, and λ determines the relative importance of good reconstructions and sparsity.

Generative model for Equation (1)



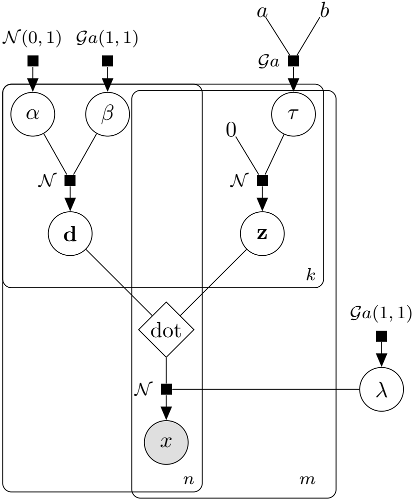
- Based on Yang et al. (8).

Generative model for Equation (1)



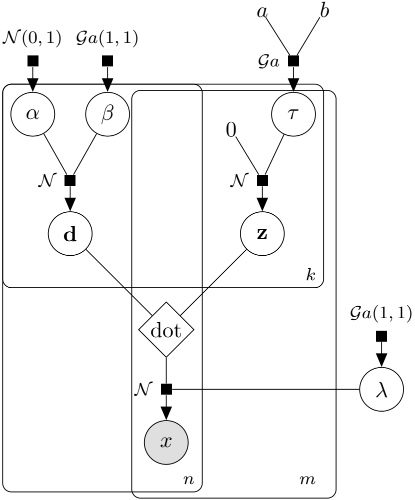
- Based on Yang et al. (8).
- Gamma prior over β, which allows the dictionary atoms to be automatically scaled

Generative model for Equation (1)



- Based on Yang et al. (8).
- Gamma prior over β, which allows the dictionary atoms to be automatically scaled
- additional level of hierarchy through the variables α on the means of the dictionary components, which can aid online learning

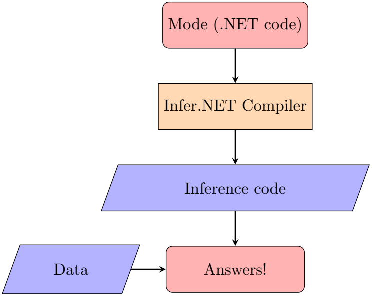
Generative model for Equation (1)



- Based on Yang et al. (8).
- Gamma prior over β, which allows the dictionary atoms to be automatically scaled
- additional level of hierarchy through the variables α on the means of the dictionary components, which can aid online learning
- a, b control sparsity:

  $\mathcal{G}a(1,1) \rightarrow$ non-sparse
  $\mathcal{G}a(0.5, 10^-6) \rightarrow$ sparse

Infer.NET



- Framework for running Bayesian inference in graphical models[a]
- Experiments used Mono[b], running on OS-X and Linux
- Inference engine: Variational Message Passing (VMP), deterministic approximation algorithm

[a] http://research.microsoft.com/infernet
[b] http://www.mono-project.com/

Modelling code

```
public void BDL(double[,] x, int numBases, double a, double b)
{
  var basis = new Range(numBases);
  var signal = new Range(x.GetLength(0));
  var sample = new Range(x.GetLength(1));

  var noisePrecision = Variable.GammaFromShapeAndRate(1, 1);
  var coefficientPrecisions = Variable.Array<double>(signal, basis);
  var dictionaryMeans = Variable.Array<double>(basis, sample);
  var dictionaryPrecisions = Variable.Array<double>(basis, sample);
  var coefficients = Variable.Array<double>(signal, basis);
  var dictionary = Variable.Array<double>(basis, sample);
  var signals = Variable.Array<double>(signal, sample);

  dictionaryMeans[basis][sample] = Variable.GaussianFromMeanAndVariance(0, 1).ForEac
  dictionaryPrecisions[basis, sample] = Variable.GammaFromShapeAndRate(1, 1).ForEach
  coefficientPrecisions[signal, basis] = Variable.GammaFromShapeAndRate(a, b).ForEac
  coefficients[signal, basis] = Variable.GaussianFromMeanAndPrecision(0, coefficient
  dictionary[basis, sample] = Variable.GaussianFromMeanAndPrecision(
        dictionaryMeans[basis, sample], dictionaryPrecisions[basis, sample]);

  dictionary[basis, sample].InitialiseTo(dictionaryPriors[basis, sample]);

  var cleanSignals = Variable.MatrixMultiply(coefficients, dictionary).Named("clean"
  signals[signal, sample] = Variable.GaussianFromMeanAndPrecision(cleanSignals[signa

  signals.ObservedValue = x;
}
```

Datasets

HAR dataset Anguita et al. (1)

- Publicly available activity recognition dataset[a]
- Smart-phone on the waist, with tri-axial accelerometer, 50 Hz
- Annotation was done using video-recordings.
- Activities: Lie, Stand, Walk, Sit, Ascend stairs, Descend stairs

[a] https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

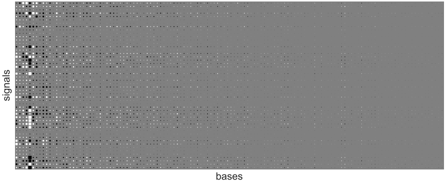
SPHERE challenge dataset Twomey et al. (7)

- collected by SPHERE project and made public as a challenge[a]
- Tri-axial accelerometer on dominant wrist, 20 Hz, range ± 8 g
- Activites: Lie, Stand, Walk

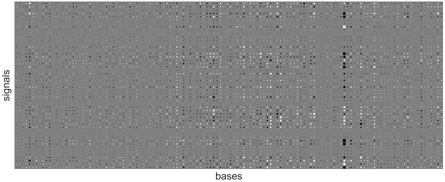
[a] http://irc-sphere.ac.uk/sphere-challenge/home

Sparsity

Non-sparse priors ($\mathcal{G}a(1,1)$) **average sparsity 0.84**

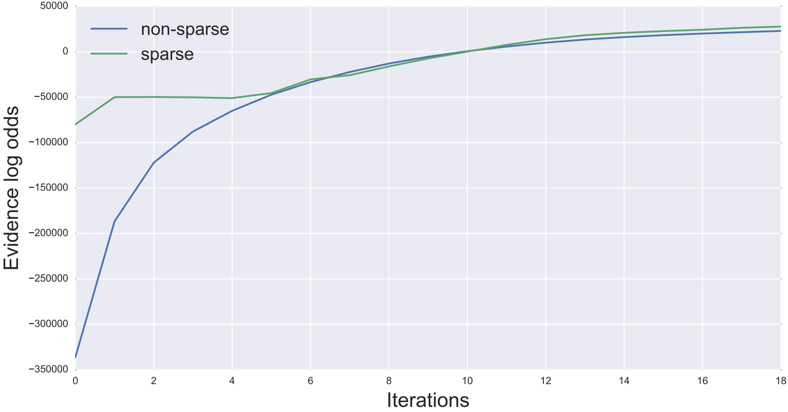


Sparse priors ($\mathcal{G}a(0.5, 10^{-6})$) **average sparsity 0.96**

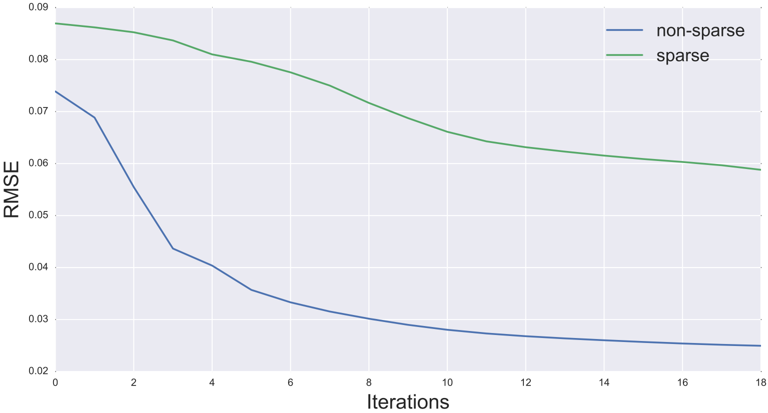


Convergence of VMP

- Model evidence converges monotonically to a local maximum



Convergence of reconstruction error



Learnt Dictionary

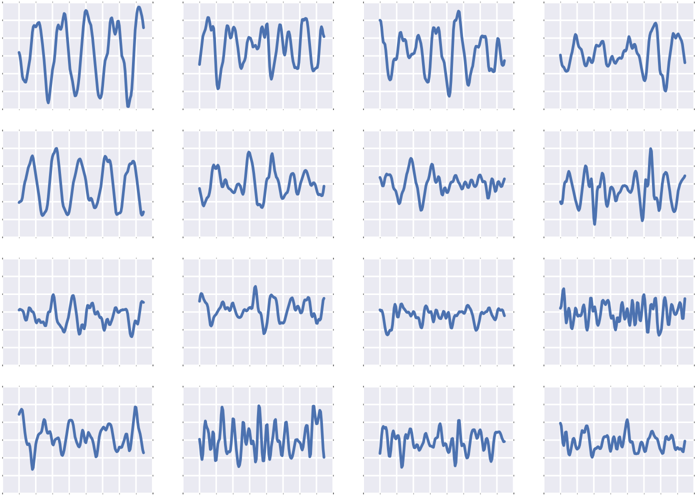
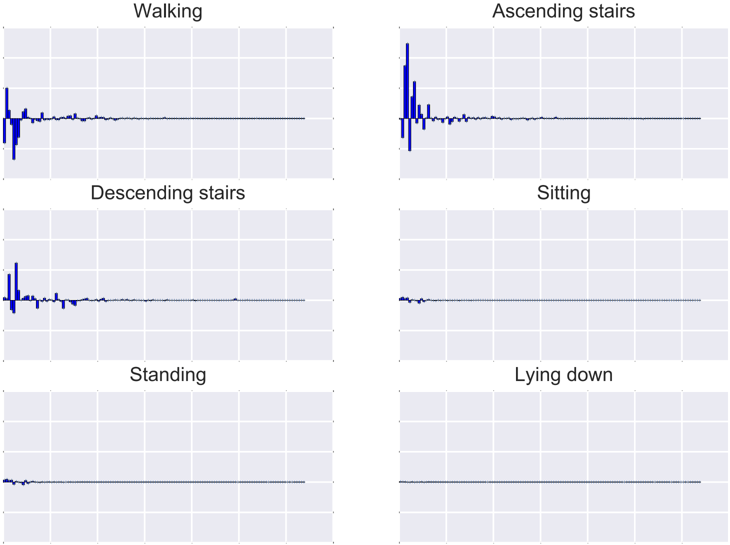
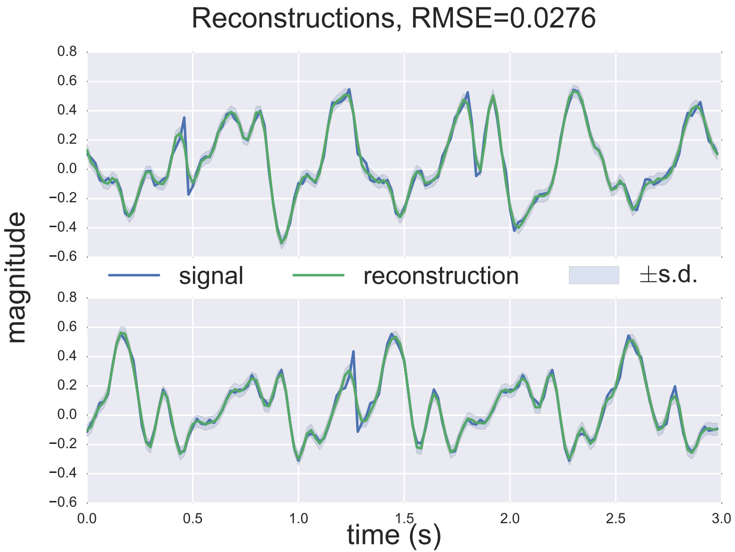


Figure: 16 example bases from the dictionary of 128 bases inferred by BDL using the non-sparse priors on HAR dataset

Coefficients



Reconstructions

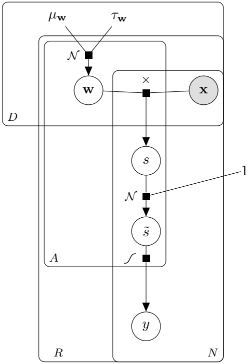


Reconstruction Performance

	SPAMS		BDL Sparse		BDL	
Bases	RMSE	Sparsity	RMSE	Sparsity	RMSE	Sparsity
64	0.0480	0.71	0.0519	0.88	**0.0293**	0.62
128	0.0457	0.84	0.0400	0.94	**0.0276**	0.84
256	0.0438	0.92	0.0316	0.99	**0.0288**	0.93
512	0.0423	0.96	**0.0224**	0.99	0.0231	0.96

Classification

- Multi-class Bayes Point Machine (BPM) Herbrich et al. (5), also using Infer.NET
- Assumptions:
  - Feature values **x** are always fully observed
  - The order of instances does not matter
  - The predictive distribution is a linear discriminant of the form: $p(y_i|\mathbf{x}_i, \mathbf{w}) = p(y_i|s_i = \mathbf{w}'\mathbf{x}_i)$, where **w** are the weights and s_i is the score for instance i



Activity Recognition Results

- Use coefficients as features of classification algorithm
- 64 bases + bias feature
- Metric: per-class 1 vs rest area under ROC curve

HAR dataset

Activity	BDL sparse	BDL	SPAMS
Walking	0.73	0.83	0.88
Ascending stairs	0.63	0.60	0.83
Descending stairs	0.61	0.34	0.82
Sitting	0.74	0.72	0.89
Standing	0.51	0.43	0.98
Lying down	0.95	0.95	0.95
Average	0.70	0.65	0.89

SPHERE dataset

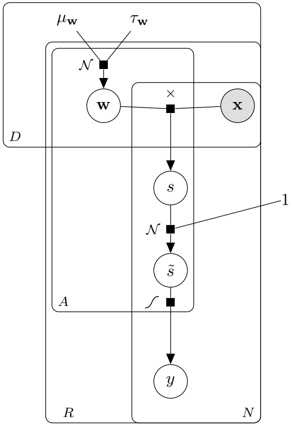
Activity	BDL sparse	BDL	SPAMS
Walking	0.55	0.51	0.46
Standing	0.69	0.62	0.44
Lying down	0.86	0.85	0.46
Average	0.70	0.66	0.45

Transfer Learning

- Source: $D_S = \{(\mathbf{x}_i, y_i)_{i=1}^{m_S}\}$
- Target: $D_T = \{(\mathbf{x}_i, y_i)_{i=1}^{m_T}\}$
- Assumptions:
  - labels available for source domain
  - labels can be acquired for target domain, but costly

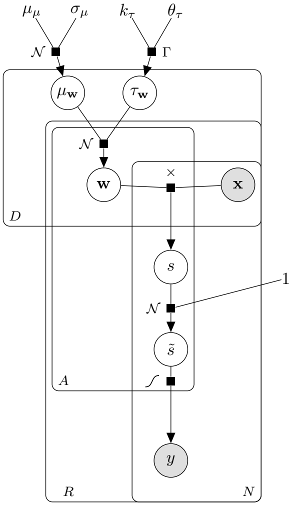
Hierarchical Model

Multiclass Bayes Point Machine (BPM)



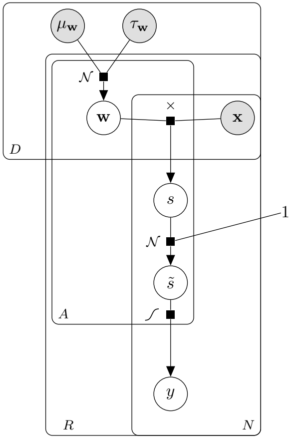
Hierarchical Model

Community: Learn community posteriors



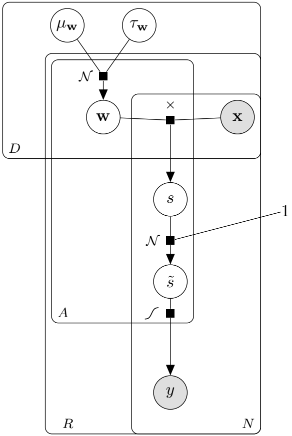
Hierarchical Model

Transfer: weight means and precisions replaced by community posteriors from source



Hierarchical Model

Personalisation: smoothly evolve from generic to custom predictions



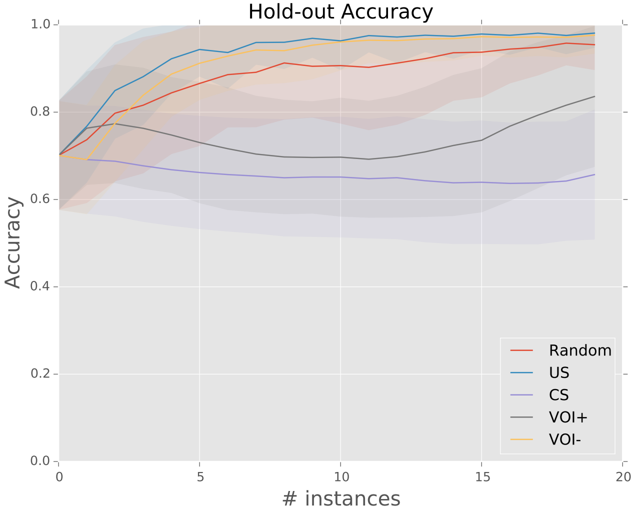
Active Learning

- Baseline: random (= online learning)
- Uncertainty Sampling
  Nearest to decision boundary, fails in case of "corrupted" data
- Certainty Sampling
  Furthest from decision boundary. Normally nonsensical but solves "corruption" issue
- Value of Information (VOI) (6)
  Decision theoretic measure that uses $p(y = 1)$ vs $p(y = 0)$ (relative risk). Expensive to compute.

Datasets

- Source: (1) 30 subjects, Smartphone, 50Hz, Video annotations
- Target: (9) 14 subjects, MotionNode, 100Hz, Observer annotations
- Classes: Walking upstairs vs. Walking downstairs

Results


Hold-out Accuracy
Accuracy
instances
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
Random
US
CS
VOI+
VOI-

Conclusions and Further Work

- Improved methods for Bayesian Dictionary Learning
- Sparsity does not need to be enforced
- Application to accelerometer signals for Activity Recognition
- Hierarchical extension to BPM for transfer learning
- Combination of Active & Transfer learning
  $\longrightarrow$ fast learning on target dataset
- Uncertainty Sampling good enough

Conclusions and Further Work

- Improved methods for Bayesian Dictionary Learning
- Sparsity does not need to be enforced
- Application to accelerometer signals for Activity Recognition
- Hierarchical extension to BPM for transfer learning
- Combination of Active & Transfer learning
  $\longrightarrow$ fast learning on target dataset
- Uncertainty Sampling good enough
- Further Work:
  - Incorporate classifier directly into the model

Conclusions and Further Work

- Improved methods for Bayesian Dictionary Learning
- Sparsity does not need to be enforced
- Application to accelerometer signals for Activity Recognition
- Hierarchical extension to BPM for transfer learning
- Combination of Active & Transfer learning
  $\longrightarrow$ fast learning on target dataset
- Uncertainty Sampling good enough
- Further Work:
  - Incorporate classifier directly into the model
- **Source code:**

  https://github.com/IRC-SPHERE/bayesian-dictionary-learning

  https://github.com/IRC-SPHERE/ActiveTransfer

Topic Models for ADL

- Supervised Learning is hard (expensive labels etc). We've seen active and transfer learning. Can we go completely unsupervised?
- Topic models: probabilistic models for discovering the latent structures in (text) documents
- Occurrence of an activity $\longrightarrow$ segment sensor data. Activities as latent structure
- But: need to segment the sensor data first
- Solution: Unsupervised segmentation (2)

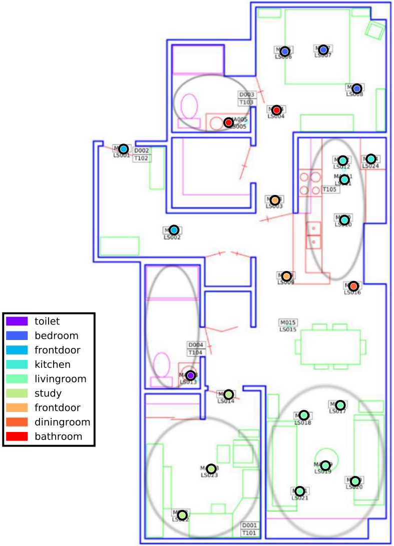
Smart Home Environmental Sensors

Timestamp	Sensor	Reading
2013-04-01 00:04:09.340911	M007	ON
2013-04-01 00:04:10.485392	M007	OFF
2013-04-01 00:56:31.879063	T106	24
2013-04-01 01:13:53.616434	BATV104 3070	
...	...	...
2013-04-01 02:45:47.215554	M006	OFF

- Motion, door, and light sensors, provide binary readings, whereas temperature sensors, battery sensors provide continuous values

Dataset	# Activities	# Binary Sensors	Duration (days)	# Residents
hh122	32	24	30	1
hh120	32	24	64	1
milan	15	31	31	1+pet

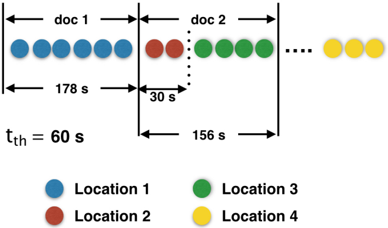
Sensor Locations



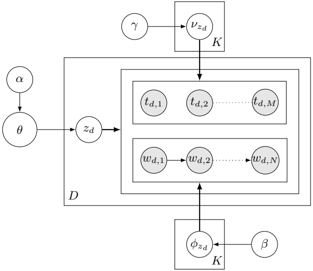
Sensor Words and Sensor Documents

- id + reading $\longrightarrow$ "word"
- For M binary sensors we obtain a vocabulary with $2M$ unique words
- locations highly correlated with activities
  - change of location strong signal of a switch between activities
  - but some activities involve motion
  - require segments to be of a minimum duration, expressed by the time threshold t_{th}

Sensor ID	Sensor Reading	Word ID
M007	ON	1
M007	OFF	2
...	...	...
D004	OPEN	48



Model



Two independent word sequences composing one document:

1. a sequence of unigrams which are independently drawn;
2. a sequence of bigrams which are represented by a Markov chain.

Unlike conventional topic models, topics are drawn for documents rather than words

Algorithm 2: Generative Processes of ADL^{TM}

Draw a $\theta \sim Dir(\alpha)$;
for $d = 1$ *to* D **do**
  Draw a topic $\mathbf{z}_d \sim Multi(\theta)$;
  Draw a $\phi_{z_d} \sim Dir(\beta)$;
  Draw a $\nu_{z_d} \sim Dir(\gamma)$;
  for $n = 1$ *to* N **do**
    Draw a unigram $t_{d,n}|z_d \sim Mult(\nu_{z_d})$;
    if $n > 1$ **then**
      Draw a bigram $w_{d,n}|w_{d,n-1}, z_d \sim Mult(\phi_{z_d})$;
  end
end

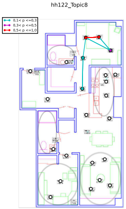
Gibbs Sampling for ADL^{TM}:

$$P(z_d = k|\mathbf{z}_{-d}, \mathbf{w}, \mathbf{t}) \propto$$

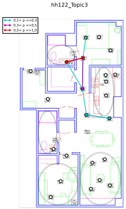
$$P(z_d = k|\mathbf{z}_{-d}) \prod_{n=1}^{M} P(t_{d,n}|z_d = k, \mathbf{z}_{-d}, \mathbf{t}_{-d})$$

$$\times \prod_{n=2}^{N} P(w_{d,n}|w_{d,n-1}, z_d = k, \mathbf{z}_{-d}, \mathbf{w}_{-d})$$

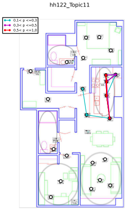
Topics Discovered



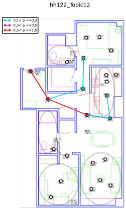
topic 8



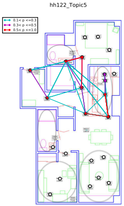
topic 3



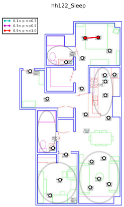
topic 11



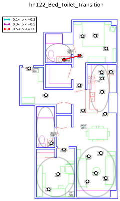
topic 12



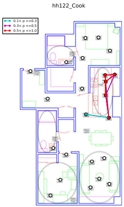
topic 5



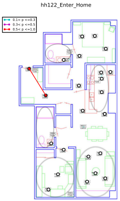
Sleep



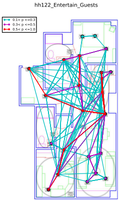
Bed-to-Toilet



Cook



Enter Home



Entertain Guests

Performance evaluation

- Tested on several CASAS datasets which were collected in real smart homes and *partially annotated*
- Compared with randomly assigned topics and other two popular topic models: Latent Dirichlet Allocation (LDA) and Bigram Topic Model (BTM)
- Clustering performance
  - Use Fowlkes-Mallows index, a variant (geometric mean) of the F_1 score adapted for clustering
  - F_1: Harmonic mean (for rates) of precision and recall $H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$

	hh122	hh120	milan
Random topics	0.0798	0.0969	0.1193
BTM	0.1515	0.1988	0.3225
LDA	0.3268	0.3486	0.5634
ADL^{TM}	0.3362	0.4072	0.6190

Performance of ADL^{TM}

- Segmentation Performance
  - **Segmentation Error**:
    $Err_s = \sum_i^{D_s} E_i / N_{dp}$, where D_s: # of generated segments, N_{dp}: # of data points, E_i: # of points in segment i who do not belong to the dominant activity of this segment:
    $E_i = N_i - \sum_{j=1}^{N_i} I(a_{ij} = m), m = \arg\max_k(\sum_{j=1}^{N_i} I(a_{ij} = k))$ where N_i: # of data points in segment i, a_{ij} is the annotated activity of point j in segment i, $I(x)$ is the indicator function
  - **Fragment Ratio**:
    $R_{fr} = D_s / D_a$, where D_a: # of occurrences of activities in the evaluated data. R_{fr}: average # of segments in one occurrence of an activity.

	hh122		hh120		milan	
	Err_s	R_{fr}	Err_s	R_{fr}	Err_s	R_{fr}
Documents	0.0197	1.596	0.0404	2.084	0.0342	1.563
LDA	0.0541	1.178	0.0531	1.554	0.0409	1.157
BTM	0.0619	1.174	0.0568	1.853	0.0428	1.394
ADL^{TM}	0.0512	1.102	0.0516	1.364	0.0406	1.149

Scope & Limitations

- +ves
  - Does not need large amount of annotated data or expensive estimation of hyper-parameters
  - Easy to interpret
  - Discovered topics could be used to segment and label the raw data automatically
- -ves
  - does not discriminate temporal patterns of activities
  - assumes activities appear independently
  - can not increase the number of topics automatically while there are new activities arises incrementally

Re-cap

Goal: Activity Recognition in Smart Homes

- Noisy data
  $\longrightarrow$ methods that characterise the noise
- Missing packets
  $\longrightarrow$ robust methods
- Limited storage and transmission capacity
  $\longrightarrow$ sparse/compressive methods
- Labelled data time-consuming and costly to acquire
  $\longrightarrow$ Active Learning
- Differing training and deployment settings
  $\longrightarrow$ Transfer Learning
- Can we do without labels at all?
  $\longrightarrow$ Unsupervised Learning

Resources

- SPHERE Code Available at:
  https://github.com/IRC-SPHERE
- SPHERE Challenge Dataset
  http://irc-sphere.ac.uk/sphere-challenge/home
- Infer.NET
  http://research.microsoft.com/en-us/um/cambridge/projects/infernet/
- CASAS datasets:
  http://casas.wsu.edu/datasets/

Selected References

[1] D Anguita, A Ghio, L Oneto, X Parra, and JL Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones. In *ESANN*, 2013.

[2] Yu Chen, Tom Diethe, and Peter Flach. Adl$^{\mathbf{TM}}$: A topic model for discovery of activities of daily living in a smart home. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016.

[3] Tom Diethe, Niall Twomey, and Peter Flach. Active transfer learning for activity recognition. In *24th European Symposium on Artificial Neural Networks, ESANN 2016, Bruges, Belgium, April 27-29, 2016*, 2016.

[4] Tom Diethe, Niall Twomey, and Peter Flach. BDL. NET: Bayesian dictionary learning in Infer. NET. In *Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on*, pages 1–6. IEEE, 2016.

[5] Ralf Herbrich, Thore Graepel, and Colin Campbell. Bayes point machines. *Journal of Machine Learning Research*, 1:245–279, January 2001. URL http://research.microsoft.com/apps/pubs/default.aspx?id=65611.

[6] Ashish Kapoor, Eric Horvitz, and Sumit Basu. Selective supervision: Guiding supervised learning with decision-theoretic active learning. In *IJCAI*, volume 7, pages 877–882, 2007.

[7] Niall Twomey, Tom Diethe, Meelis Kull, Hao Song, Massimo Camplani, Sion Hannuna, Xenofon Fafoutis, Ni Zhu, Pete Woznowski, Peter Flach, and Ian Craddock. The sphere challenge: Activity recognition with multimodal sensor data. *arXiv preprint arXiv:1603.00797*, 2016.

[8] Linxiao Yang, Jun Fang, Hong Cheng, and Hongbin Li. Sparse Bayesian dictionary learning with a Gaussian hierarchical model. *CoRR*, abs/1503.02144, 2015. URL http://arxiv.org/abs/1503.02144.

[9] M. Zhang and A.A. Sawchuk. USC-HAD: A daily activity dataset for ubiquitous activity recognition using wearable sensors. In *ACM Int. Conf. on Ubiquitous Computing Workshop on Situation, Activity and Goal Awareness (SAGAware)*, 2012.

tdiethe@amazon.com