Towards occupancy inference in a smart home using energy and environmental data sources and binary classification techniques

Dr. Stelios Krinidis
Information Technologies Institute
Center for Research and Technology Hellas
krinidis@iti.gr

Outline

- Introduction – Problem definition
- Dataset presentation
- Machine learning algorithms
- Grid search
- Simulation results
- Conclusion

Introduction

The problem

The challenging problem of occupancy detection in a domestic environment is studied based on information energy consumption and environmental features.

Our approach to address the problem of occupancy detection is through Machine Learning

The most popular machine learning techniques for binary classification are utilized are tested for different sets of internal parameters so as to reach a dominant and effective classification model for occupancy detection.

Dataset

Dataset collection

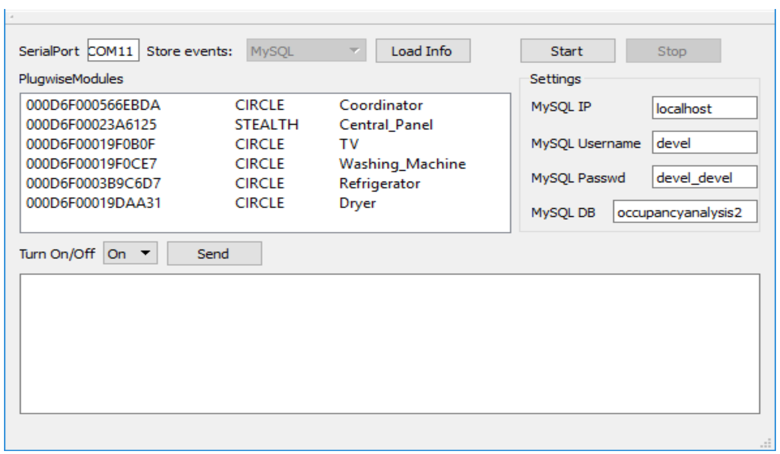
Database

- ❑ Energy consumption
- ❑ Environmental features

Training data

- ❑ Occupancy

Ground truth



Collection information application


```
{
    "version": "v1",
    "timestamp": "2016-07-21T09:00:00.000+03:00",
    "event": {
        "mValue": 0.010351999662816525,
        "dSensor": "HH_02_GHQNP.mwm_2_IBWFA",
        "readings": 4,
        "dYear": "2016"
    }
},
{
    "version": "v1",
    "timestamp": "2016-07-21T10:00:00.000+03:00",
    "event": {
        "mValue": 0.0034760001581162214,
        "dSensor": "HH_02_GHQNP.mwm_2_IBWFA",
        "readings": 4,
        "dYear": "2016"
    }
},
{
    "version": "v1",
    "timestamp": "2016-07-21T11:00:00.000+03:00",
    "event": {
        "mValue": 0.0015480000292882323,
        "dSensor": "HH_02_GHQNP.mwm_2_IBWFA",
        "readings": 4,
        "dYear": "2016"
    }
},
```


Dataset features

- Overall energy consumption
- Temperature
- Humidity
- Humidity ratio
- Luminance

Machine Leaning

- Machine Learning
  - Support Vector Machines
    - Polynomial
    - Radial Basis Function
  - Decision Tree
  - Random Forest
  - Back propagation network
  - Naïve Bayes
  - Logistic regression

- Data annotation
  - Door counter measurements
    - Classes
      - Absence
      - Presence

- Evaluations measures

		Predicted class	
		Absence	Presence
Actual Class	Absence	TP	FN
	Presence	FP	TN

- Precision
- Recall
- Accuracy
- F-measure

Grid search

Classifier	Simulation Parameters
SVM POLY	p = 2, 3, 4, 5, 6 θ = 10, 20, 30, 40
SVM RBF	C = 100, 500, 1000 σ = 0.001, 0.01, 0.1
Decision Tree	CART Algorithm
Random Forest	Number of trees = 20, 40, 60, 80, 100
Back Propagation Network	n = 100, 120, 140, 160, 180,200
Naïve Bayes	Gaussian assumption
Logistic regression	Tolerance = 1e-4 C = 1.0

- Simulation cases overall
  - 39
- Features tested
  - 5
- Ground truth
  - Yes
- Monte Carlo Iterations
  - 100
- Cross Validation
  - *Random sampling* 70%-30%

Simulation results - I

Support Vector Machines – Radial Basis Function – Polynomial kernels

SVM RBF

C	σ	Precision (%)	Recall (%)	Accuracy (%)	F_1-score (%)
100	0.001	**98.93**	27.69	72.83	43.27
100	0.01	90.91	30.77	72.83	45.98
100	0.1	86.67	40.00	75.14	54.74
500	0.001	97.86	23.53	69.94	37.93
500	0.01	95.65	31.88	72.25	47.83
500	0.1	96.55	**40.58**	**75.72**	**57.14**
1000	0.001	98.62	28.57	71.10	44.30
1000	0.01	91.67	31.43	71.10	46.81
1000	0.1	83.87	37.14	71.68	51.49

SVM Polynomial

p	θ	Precision (%)	Recall (%)	Accuracy (%)	F_1-score (%)
2	10	79.17	32.76	74.57	46.34
2	20	78.57	37.93	75.72	51.16
2	30	79.31	39.66	76.30	52.87
2	40	79.31	39.66	76.30	52.87
3	10	61.82	44.59	73.42	51.81
3	20	73.18	48.24	73.96	58.14
3	30	78.12	43.10	76.88	55.56
3	40	**80.65**	43.10	**77.46**	56.18
4	10	62.22	48.28	72.83	54.37
4	20	57.63	58.62	71.68	58.12
4	30	50.00	51.72	66.47	50.85
4	40	54.84	58.62	69.94	56.67
5	10	65.38	58.62	75.72	**61.82**
5	20	58.06	**62.07**	72.25	60.00
5	30	68.42	54.17	70.52	60.47
5	40	67.31	55.21	68.92	60.67
6	10	59.70	57.97	67.63	58.82
6	20	61.14	56.98	66.10	58.99
6	30	62.05	55.27	66.32	58.46
6	40	62.76	56.74	68.86	59.60

***highest values in bold**

Simulation results - II

Naïve Bayes – Logistic regression – Decision trees – Random forest - Back propagation network

Naïve Bayes – Logistic regression – Decision trees				
Metric	Precision (%)	Recall (%)	Accuracy (%)	F_1-score (%)
Naïve Bayes	46.83	**80.82**	53.18	59.30
Logistic regression	**97.36**	33.82	73.99	50.47
Decision Tree	68.18	66.18	**74.57**	**67.16**

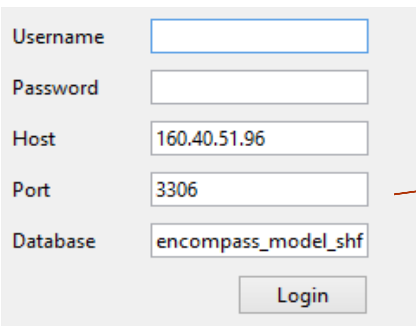
Random forest				
Estimators	Precision (%)	Recall (%)	Accuracy (%)	F_1-score (%)
20	67.21	62.12	73.99	64.75
40	72.41	62.69	76.30	67.20
60	71.43	58.82	74.57	64.52
80	**74.44**	62.36	**76.88**	68.25
100	73.33	**64.71**	75.94	**68.75**

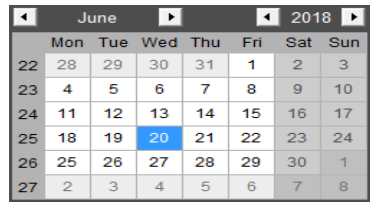
Back - propagation network				
Neurons	Precision (%)	Recall (%)	Accuracy (%)	F_1-score (%)
100	90.44	24.64	69.36	39.08
120	77.78	**30.43**	68.79	43.75
140	75.00	30.24	68.21	**43.30**
160	**94.44**	24.64	**69.36**	39.08
180	93.12	23.92	68.26	38.06

***highest values in bold**

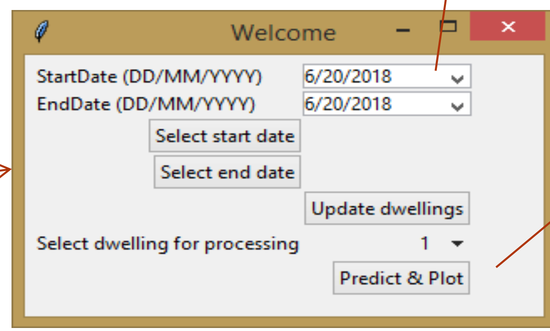
Graphical User Interface

Login

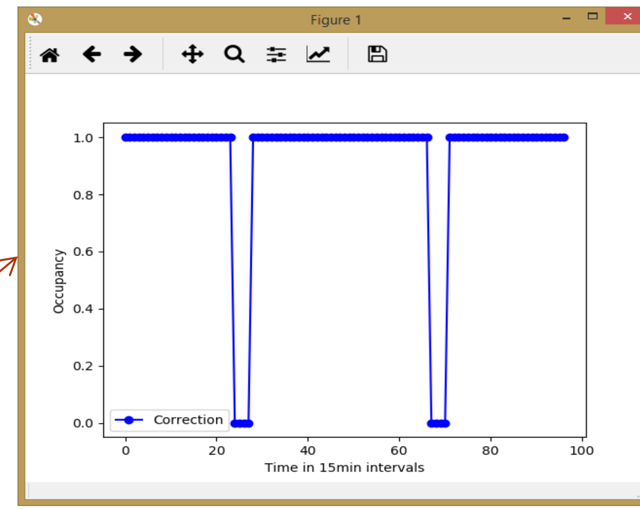




Dates selection



Daily occupancy inference



Conclusions

Simulation results overall

- **Random Forest** - (Accuracy: 76.88%, F-measure: 68.7%)
- **Decision Trees** - (Accuracy: 74.57%, F-measure: 67.16%)

Future work
- **Additional features**
- **Feature selection techniques**
- **Neural networks**

The End…

Presenter:

Dr. Stelios Krinidis

krinidis@iti.gr