**Inference from Electricity Consumption Data**

**Daniel Christopher Peric 25811460**

**Andisheh Partovi 22999825**

# Introduction

In this project, we have applied two statistical methods to the power consumption data acquired from 9 houses to determine what fraction of the year each house is occupied and whether a house is a holiday home or not.

In this report, we first introduce the system and some of it capabilities (The Chatterbot), then discuss the procedure we undertook to develop the chatterbot (Procedure) and the evaluation approach we used to test it (Evaluation). We will also discuss different dialogue phenomena we handled in the system (Handling Different language and Dialogue Phenomena) and some of the limitations of the used framework (Framework Limitations).

# The Data

One of the main limitations of this dataset is that we don’t “know” whether someone is home or not, an assumption must be made from the data. If the assumption is that when power is not being used then the person is not home, that is a bad assumption as the lowest power usage is at night when people are sleeping and most likely home. If it is true that a person uses only a little power when sleeping then we can assume that the power used when sleeping is similar to or the same as the latent or background power usage.

# Methods

We have utilized two method to determine the states of the houses. Firstly, a simple frequentist method that works with the mean and standard deviation of the data and second, a Bayesian Hidden Markov Model. In this section, we elaborate on these methods.

## Hidden Markov Model

The second method utilized was a Hidden Markov Model with the following graphical representations. Two states were assumed for the model: ‘A’ for being Away and ‘H’ for being Home. The model can become more complicated by adding extra states such as one person being home.

A straightforward approach would have been running the model for all observations and determine if the residences are home or away every half an hour. Then a majority voting system can be used to determine if they were home on that day or not. This raises the question of what constitutes being home. For our experiments, we decided to and the most likely sequence is given by the Viterbi algorithm.

### The Model

Due to the time limitations of the project, no learning algorithm was implemented for this model and the parameters of the model were set for each house according to a series of assumptions and observations that will be explained shortly. If later the Baum-Welch algorithm is implemented, the current model can serve as a good initial guess for the iterative leaning procedure. The parameters of the model are presented in the Appendix I.

1. *The Initial Probabilities:* were considered uniform for all non-holiday houses as there is no extra information to estimate a better model. Many Australians prefer to stay home on first of January and go away on vacation afterwards. For the holiday houses, the probability of being home on first of January is higher as we assumed that having a holiday house shows the family’s desire to use it during holidays.
2. *The Transitional Probabilities:* were estimated by looking at the 2013 Australian calendar. We developed two sets of transitional probabilities, one assuming that families go on holiday only twice a year during the Christmas and Easter breaks and the rest are at home, and second, that they also go away on weekends. Based on this assumptions, sequences of 365 ‘A’s and ‘H’s were created and the transitional probabilities were calculated by simply counting and normalizing the four pairs (‘AA’, ‘AH’, ‘HA’, and ‘HH’). Comparing the results of the two models also provided a good sensitivity analysis on the effects of the transitional probabilities. A same procedure was employed for holiday houses but with the inverse of the previous sequence.
3. The Emission Probabilities: were described with a Gaussian distribution CDF and the hyper parameters of the distribution were calculated using the data. The mean was the mean of and the standard deviation was the standard deviation of.

# Results and Analysis

The transitional probability matrix was estimated by looking at the calendar and assuming two different scenarios the people go on holiday twice a year. This model is generalized and we expected that it would not perform well on all houses, especially if they have a habit of going out on weekends. Developing a more specialized model for each house should improve the performance while adds more complexity.

# Appendix I – Parameters of the HMM

**Non-holiday Houses**

Model 1. Only away twice a year.

|  |  |  |
| --- | --- | --- |
| States | A | H |
| A | 0.867 | 0.133 |
| H | 0.006 | 0.994 |

Model 2. Away twice a year for a long period and also away on the weekends.

|  |  |  |
| --- | --- | --- |
| States | A | H |
| A | 0.550 | 0.450 |
| H | 0.192 | 0.808 |

House 1

|  |  |  |
| --- | --- | --- |
|  | Mean | Standard Deviation |
| Away | 103.066 | 46.729 |
| Home | 621.303 | 657.912 |