Uncovering and Leveraging Latent Structure in Energy Usage Data

Clementine Jacoby, Katelyn Gao

1 Introduction

Utility companies have millions of customers, many of them businesses. Utilities want to help their customers be more energy efficient, especially during peak usage events when they pay market rates to supply energy to meet customer demand that surpasses their own generative capacity. Furthermore, because energy efficiency is a huge public concern, third party companies and the U.S. government, are both heavily invested in helping businesses save energy.

Targeted advice can drive large scale energy efficiency. But to give good advice, utilities need to know what type of business they're talking to (e.g. you shouldn't tell restaurants to turn off their second freezers, albeit a great tip for homeowners). Similarly, giving useful feedback means comparing businesses to others of the same type (e.g. you shouldn't harangue grocery stores for using more energy than a gas stations, but you might notify them if they use more than any grocer in the state).

Unfortunately, utilities don't have the information they need to drive efficiency. Business types aren't labeled, nor are other useful properties (like opening/closing times), which would enable better comparisons, or nuanced analysis (e.g.,wasted usage when closed). Instead, utilities have detailed energy usage data - readings straight from the meter. With the advent of AMI (Advanced Metering Infrastructure) data, utilities now have access to readings taken every 15 minutes. This data has rich underlying structure that utilities could leverage in helping customers save energy. If reliably and automatically determined, this structure could help drive energy efficiency on a massive scale. The focus of our project is to uncover latent structure from raw usage data by **1) Discovering meaningful states that could be leveraged to drive efficiency 2) Predicting business types to enable fair comparison and 3) Investigating whether these two methods can be fruitfully combined.** U.S. utilities work at a huge scale, so augmenting these datasets with useful structure could save money, energy, and the world.

2 Previous Work

We found no published papers attempting to predict business type from energy usage data. However, several papers tackle the related problems of grouping customers based on dynamic usage behavior and predicting energy consumption.

Clustering customers would facilitate the implementation of targeted, dynamic rates that can greatly reduce peak loads by manipulating consumer behavior [1]. Flath et al. [1] use k-means clustering of energy consumption data to segment customers into groups; they also discuss how to use the resulting clusters to design group-specific rates that reduce price discrimination and peak loads.

Utility companies want to predict energy consumption for the same reason they want to predict business type - to increase efficiency. If electric utilities can correctly predict usage, they would be able to better plan production throughout the day and distribute energy among grids as to decrease energy loss overall. Indeed, they may even be able to foresee and prevent failures caused by overloads. Dong et al. [2] apply Support Vector Machine regression to predict building energy consumption in tropical climates. Tso and Yau [3] compare the performances of regression, decision trees, and neural networks for predicting household energy usage in Hong Kong.

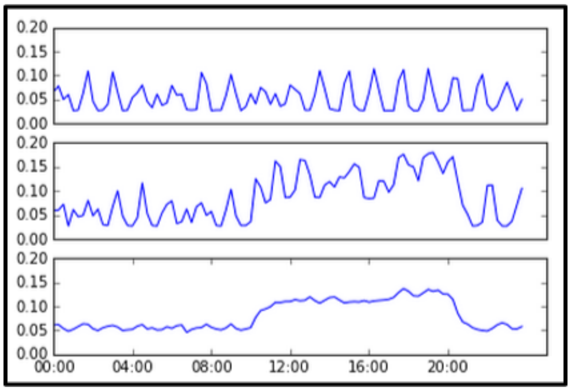

Figure 1: A Restaurant

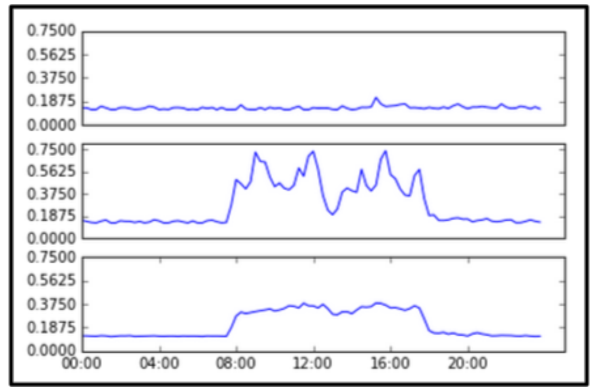

Figure 2: A Dentist

3 Problem Statement: A Threefold Quest

1) What can time-series usage data tell us about businesses? Can HMMs uncover meaningful structure in this unsupervised problem? 2) Can we distinguish a strip club from an elementary school (or restaurants from dentists, in this case) just by knowing how much energy they use, and when? More generally, can we classify businesses by type using only usage data and clever feature engineering? 3) Can we use the structure discovered in the unsupervised problem to improve our classification results?

4 Data

We obtained daily electricity consumption data for about two thousand businesses in California during January 2014. This AMI datset contains usage readings for each 15-minute interval of the day measured in kilowatt-hours. We also have a labeled dataset, collected by a manual process, containing business types for hundreds of businesses. We separate the dentists and restaurants, merge them with our AMI dataset, and remove businesses without 31 days of complete usage readings. For each business, we compute the average usage reading at each interval, obtaining a time series of length 96 that summarizes the progression of energy consumption over a day. Our final data set consists of 983 time series for dentists and 886 for restaurants, all length 96.

Figure 1 plots the energy usage for an example restaurant. The top two panels show individual days. The third shows a monthly average. These three plots summarize why this is, at once, both a difficult and a hopeful problem. The first plot exhibits no clear signal - just near-uniform noise throughout the day. The second plot is also noisy, but a signal begins to emerge - you can see when the business is open. In the third plot, this signal clearly emerges in the aggregate. Figure 2 contains the same plots for an example dentist, and we observe the same phenomenon.

5 Baseline and Oracle

There's no groundtruth in our unsupervised latent strcuture problem; we have to investigate the states by hand to see if they're reasonable. But for classification, we have good success metrics:

Baseline Algorithm: As a baseline, we run logistic regression on the 96 usage features. We frame the problem as a two-class classification problem where the labels are 'dentist' and 'restaurant'.

- Features: The usage features are the time series data, quarter-hourly measurements of usage for a business measured in kilowatt hours. For example usage_value_2200 = .363 means that this business used .363 kwh between 10 and 10:15 pm.
- Our baseline achieves 79% accuracy: This suggests that there is room for us to improve.

State	p(s)	Mean	Std. Dev.
1	0.463	0.237	0.188
2	0.455	1.143	0.399
3	0.082	3.956	2.842

Table 1: Rest. Initial and Emission Distributions

State	1	2	3
1	0.992	0.008	0
2	0.005	0.988	0.007
3	0	0.014	0.986

Table 2: Rest. Transition Probabilities

Oracle: An oracle would have access to business types for all util_service_point_ids (the unique identifiers in our dataset). These could be found by surveying businesses or by looking up business names by uspid, Googling those names, and categorizing the business manually. Both processes are expensive and manual.

6 Latent Structure: Model

For each business type, we model the energy usage over a day as a Hidden Markov Model as described below. If successful, the hidden states learned and the maximum likelihood sequences of states will help us distinguish between restaurants and dentists.

- **Hidden states**: Intuitively, these could correspond to opening and closing times or peak hours, which all have differing levels of energy usage. We only input the number of states.
- **Observations**: Usage readings for each business, averaged over a month. Each observation is a time series of length 96, containing measurements of energy usage for a given business in kilowatt hours for every fifteen minutes of a typical day
- **Transition probabilities**: The likelihood of staying in the same state and going to another state. For example, they could model the likelihood of staying closed versus opening.
- **Emission probabilities**: Given the state of a business, what is the probability of it spending a given quantity of energy? With this information, we may also deduce what real-world situations the states actually represent.

6.1 Latent Structure: Two Class Results

In our search for latent structure, we hope to learn meaningful states and posterior state sequences by modeling daily energy consumption as a Hidden Markov Model. Using the EM algorithm, we fit HMMs with Gaussian emission distributions separately on the observations from restaurants and from dentists. We experiment with different numbers of states, between 2 and 4 inclusive, and we report results learned from the 3-state HMMs. The 3-state HMMs were chosen because the 2-state HMMs often did not give meaningful posterior maximum likelihood state sequences and the 4-state HMMs did not appear to discover any latent structure not shown in the simpler 3-state models.

Tables 1 and 2 contain the parameters of the fitted HMM for restaurants. The first column of Table 1 contains the initial probability distributions of the states. The other columns are the parameters of the emission distributions of the states, and Table 2 shows the transition probabilities.

Figure 3 plots the energy usages of three restaurants and shows their posterior maximum likelihood state sequences as computed by the Viterbi algorithm [4]. Blue lines represent the times when the posterior ML state for a business is 1. Red lines represent the times when the posterior ML state is 2, and green lines represent times when the posterior ML state is 3. The paths seem reasonable; one restaurant, maybe a cafe, remains in a low energy usage state all day. The others transition to a higher energy usage state during mealtimes, from 9 am to 11 pm.

Tables 3 and 4 contain the estimated parameters of the dentist HMM, formatted as before.

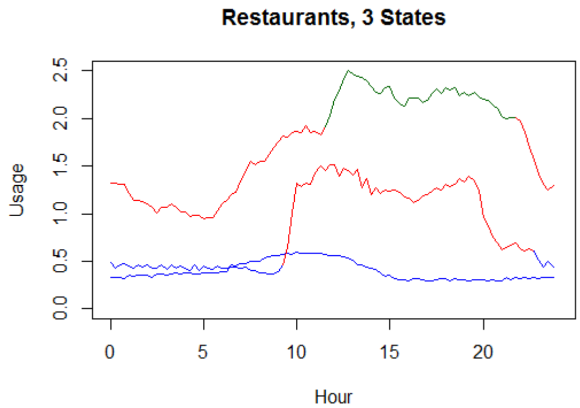

Figure 3: Posterior State Sequences

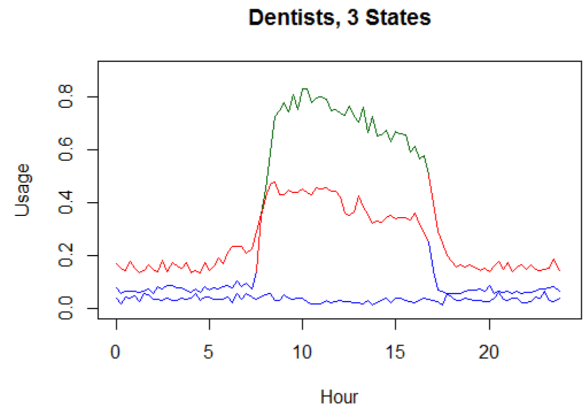

Figure 4: Posterior State Sequences

State	p(s)	Mean	Std. Dev.
1	0.538	0.071	0.045
2	0.402	0.264	0.084
3	0.060	0.890	0.443

Table 3: Dentist Initial and Emission Distributions

State	1	2	3
1	0.985	0.014	0.001
2	0.015	0.964	0.021
3	0	0.024	0.976

Table 4: Dentist Transition Probabilities

Figure 4 is the same plot as Figure 3, but for three dentists. As expected, dentists use more energy during business hours, between 8 am and 5 pm. We may conclude that the HMMs found meaningful states, infomation that could be leveraged to infer opening and closing hours for fair comparison, or used by utilities directly to manage the grid (e.g. by distributing more energy to business districts in the morning and shopping districts in the evening). Encouraged by these results, we begin to explore avenues for leveraging this uncovered structure to bolster our classification efforts.

7 HMMs for Classification

Next, we built a Bayesian classifier for whether a business is a restaurant or dentist. This begins to bridge the gap between our two approaches: we can compare the accuracy of this classifier to our baseline. Given training data from restaurants, we can use the EM algorithm to fit a HMM with 3 states and Gaussian emission distributions. We can do the same for dentists. Now, suppose that the prior probability over the types is uniform, that HMMs are the correct generative models for energy consumption time series, and that the estimated maximum likelihood parameters for both restaurants and dentists are the true values. Then, for an observed energy usage time series x, we have

$$\begin{aligned} P(dentist \mid x) \propto P(x, dentist) &= P(x, \hat{\theta}_D, dentist) = P(x \mid \hat{\theta}_D, dentist) P(\hat{\theta}_D, dentist) \\ &= 0.5 P(x \mid \hat{\theta}_D, dentist) \end{aligned} \tag{1}$$

where $\hat{\theta}_D$ are the estimated parameters of the dentist HMM. A similar expression holds for $P(restaurant \mid x)$. Therefore, we predict dentist if the likelihood of x under the fitted HMM for dentists is higher than the likelihood of x under the fitted HMM for restaurants, and restaurant otherwise.

To estimate the error rate of this classifier, we split our data into a training set and test set. The first 295 restaurants and dentists were put in the training set, and the next 591 in the test set. (The businesses were in no special order.) We carried out the procedure described above, and applied the resulting classifier

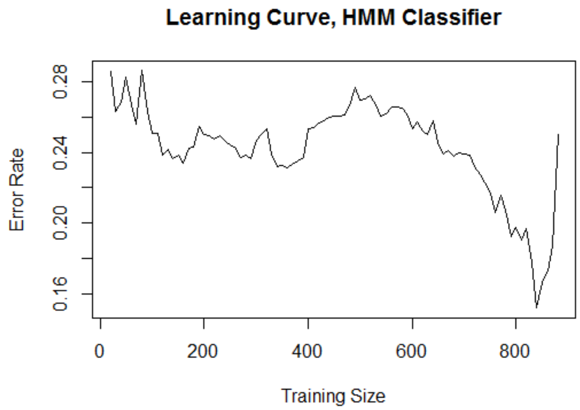

Figure 5

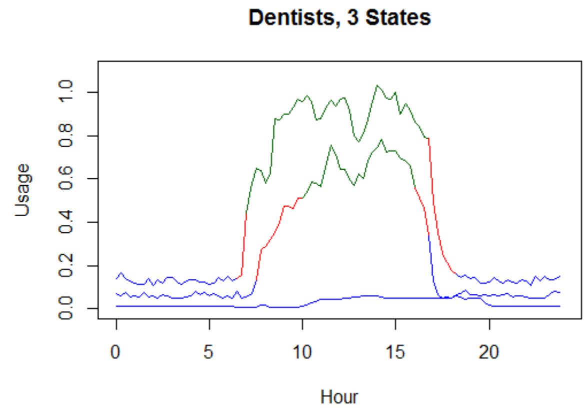

Figure 6: Posterior State Sequences in July

to the test set. Out of 1182 test observations, there were 284 errors, giving an accuracy of 76%, below that of the baseline logistic regression. This is not surprising, because states convey strictly less information than energy usage measurements.

Looking at specific observations, it seems that businesses with similar levels of energy usage throughout the day were the hardest to classify correctly. It is expected, since in those cases there would be no special structure in the dynamics of energy usage for the HMM to exploit.

We estimated the learning curve of the classifier by varying the number of restaurants/dentists in the training set; it is shown in Figure 5. It appears that using anywhere between 200 and 600 restaurants and dentists in the training set would lead to an error rate of approximately 24%.

8 Classification: Better Features for Business Type Prediction

Here, we return to the classification problem. Although HMMs are not an obvious first step in type classification, now that we know that our HMMs learn meaningful states, we want to try using the HMM output to improve classification accuracy. To test this, we use the HMM output as features in our logistic regression algorithm. We add these features (probability of being a dentist and/or the label computed by the HMM classifier) in addition to the usage features.

To improve on our baseline accuracy, we run a number of additional experiments: 1) Normalizing by baseline usage for a business, which surprisingly hurts accuracy 2) Adding the daily sum as a feature, a surrogate for normalization, which also hurts accuracy 3) Disaggregating weekday and weekend usage data, our most powerful feature of all. Features 1 and 2 are later improved by standardization, discussed below. The results of these experiments are summarized in section 9.3. Further classification improvements are discussed in section 10.3.

9 Classification: Extension to Summer Usage Data

Motivated by California's mild winter climate, we were moved to investigate whether we could improve classification accuracy by using Summer data, where a stronger or more variable signals might emerge. Here, we repeat our analysis with energy consumption data for 893 restaurants and 985 dentists during July 2014. We suspect that summer energy consumption may differ by pattern or quantity from winter consumption (e.g. restaurants may use more energy in the summer if more people go out to eat; all businesses will use more energy for air conditioning), which may aid classification. Our July data held the format of the January data, and we carried out the same pre-processing.

9.1 Latent Structure: Extension to Summer Usage Data

As in Section 6, we fit 3-state HMMs with Gaussian emission distributions on the observations from restaurants and the observations from dentists. For the most part, the conclusions about the dynamics of energy usage and hidden states are quite similar to the results from January.

However, the usage levels are higher. The highest energy usage state for restaurants now has a mean of 4.869 instead of 3.956, and the highest energy usage state for dentists now has a mean of 1.203 instead of 0.890. In Figure 6, we show the energy usages and posterior maximum likelihood state sequences for the same dentists as Figure 4. It suggests that in July, dentists tend to go to higher energy states more often and also stay there for a longer part of the day. The same is true for restaurants as well.

9.2 HMM Classification: Extension to Summer Usage Data

We estimated the error rate of the HMM classifier described in Section 7 on the July data. Again, we divided the restaurants and dentists between a training set and a test set, where the numbers of restaurants and dentists in the test set are equal. After fitting two separate 3-state HMMs with Gaussian emission distributions on the restaurants and dentists in the training set, we use the estimated parameters to construct a classifier that predicts whether each observation in the test set is a restaurant or dentist. This time, we got 77% accuracy, slightly higher that of the January data.

9.3 Classification: Better Data and Better Features United

Here we summarize the results of running LR on the unnormalized and normalized usage data for January and July. To normalize, we divide each reading by the sum of the day's usage.

Figure 7 shows accuracies for unnormalized data. We achieved higher accuracy on the July data than on the January dataset across all feature sets. This is reasonable since January is a mild month compared to July (in CA), as mentioned. Another universal trend was that our best accuracies (.901% for July, .893% for January) were obtained using weekend features, which swamp all other features in our experiments. More details on all feature sets are included below.

Classification: Experimental Feature Sets

- **just_usage**: Only the time series usage data, 96 features representing quarter-hourly readings.
- **usage_sum**: In addition to the usage features, incorporates the sum of the daily usage.
- **just_prob**: The HMM probability (probability of being a dentist computed by the HMM classifier) alone. Equivalent to running our HMM classifier on the observations (25% error rate).
- **usage_prob**: The HMM probability is added to the usage as a 97th feature. On our original dataset (January) this lead to small gains (.04%) but in July, when energy usage is more extreme due to higher ambient temperatures, the subtlety of the HMM feature is swamped by the usage features and the accuracy with and without the HMM probability is the same (88.4%).
- **usage_label**: As above, the HMM label (label output by the HMM classifier) is added to the usage as a 97th feature. Again, this led to small gains on the original dataset (.02%). The gains were a bit smaller, which is reasonable, given that the probability from the HMM represents strictly more information than the label. Again, this feature was swamped by usage in July.
- **usage_prob_label**: Includes both the HMM probability and label as 97th and 98th features.

The last four feature sets incorporate 'weekend features' (denoted _w on the x axis). We add weekend features (96 features describing weekend usage, separated from weekday usage) to the feature sets described above. In each case, adding weekend features improves the feature set, leading to an overall high of 89% for January, and 90% for July.

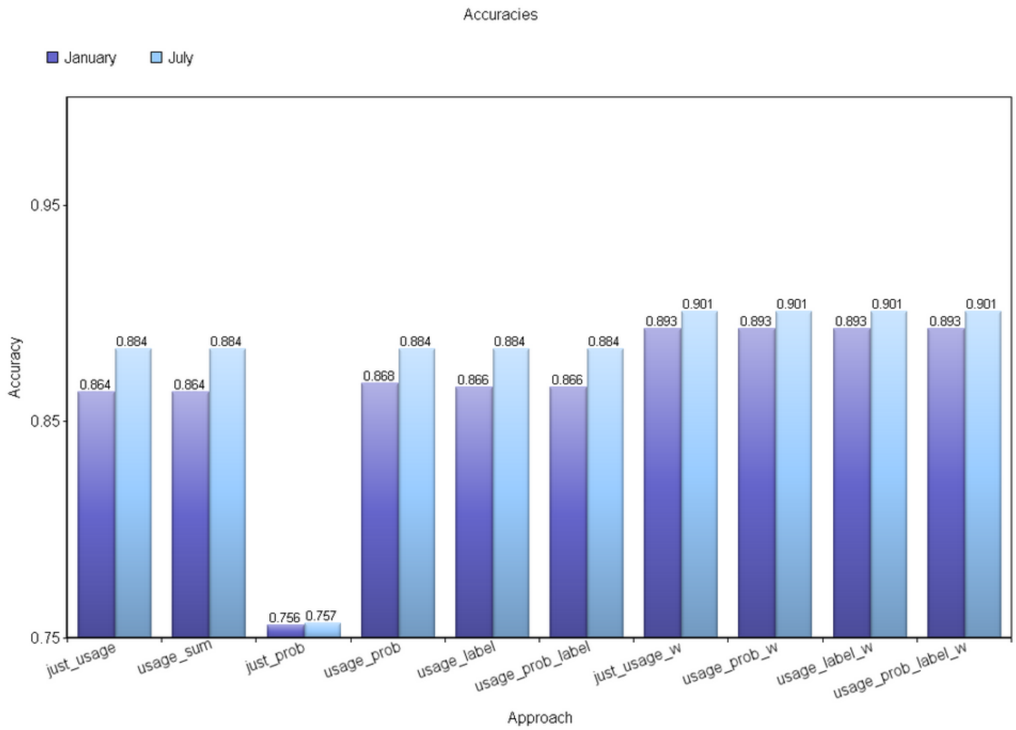

Figure 7: Unnormalized

State	p(s)	Mean	Std. Dev.
1	0.293	0.052	0.051
2	0.435	0.766	0.470
3	0.272	5.530	4.296

Table 5: School Initial and Emission Distributions

State	1	2	3
1	0.992	0.008	0
2	0.005	0.989	0.006
3	0	0.008	0.992

Table 6: School Transition Probabilities

The accuracies for the normalized data experiments were much more variable, but follow the same trends observed above - we achieve higher accuracy on July data, and the weekend feature sets achieve the same accuracies as on the unnormalized data. We achieve 80% and 85% accuracy on January and July (vs. 86% and 88% achieved before normalization). In this case, the HMM features hurt accuracy. We investigate this problem, and cure the discrepancy using standardization in section 10.3.

10 The Multiclass Case: Extension to Three Businesses Types

Moving toward the real-world situation where there are many business types, we next consider predicting whether a business is a restaurant, dentist, or school given energy usage time series data. The restaurant and dentist data is the same as in Sections 6, 7, and 8. Now, we also have daily consumption data for about 800 elementary and secondary schools in California during January 2014. We pre-process it the same way as described in Section 4 to obtain 773 time series of length 96.

10.1 Latent Structure: Multiclass Results

Following Section 6.1, we fit a 3-state HMM with Gaussian emission distributions on the school data using the EM algorithm. Tables 5 and 6 contain the fitted parameters, organized as previously. Figure 8 is the same as Figure 3, but for three schools.

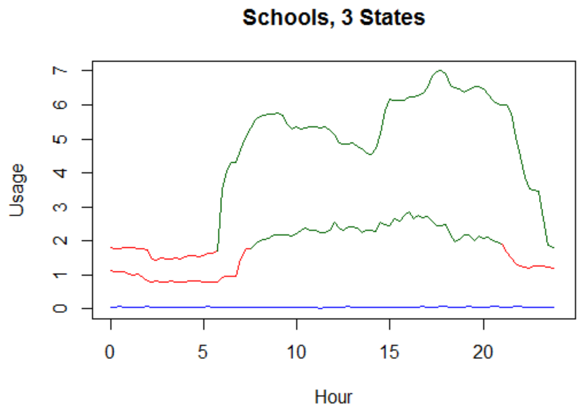

Figure 8: Posterior State Sequences

It seems that schools use more energy than restaurants and dentists, which is unsurprising given that schools house hundreds or thousands of students each weekday. The highest energy state has a mean of 5.53 (versus 3.956 and 0.89), and the initial distribution of states puts more weight on states 2 and 3. As with restaurants and dentists, the posterior maximum likelihood state sequences make sense; schools move to high energy usage states between 7 am and 8 pm, during school hours and extracurricular activities.

10.2 HMM Classification: Multiclass Results

We slightly modify the classifier described in Section 7 to obtain a classifier for this 3-class case. Making the same assumptions as before and by the same calculations, we have $P(dentist \mid x) \propto P(x \mid \hat{\theta}_D, dentist)/3$. A similar equation holds for $P(restaurant \mid x)$ and $P(school \mid x)$. Therefore, after computing the likelihood of x under all three fitted HMMs, we predict dentist if the likelihood of x is highest under the fitted HMM for restaurants; we predict restaurant if the likelihood of x is highest under the fitted HMM for restaurants, and we predict school otherwise.

To estimate the error rate of this classifier, we split the restaurants, dentists, and schools into a training set and a test set, making sure that the test set contains equal numbers of the three types. After fitting separate 3-state HMMs with Gaussian emission distributions on data belonging to each type of business in the training set, we used those estimated parameters to build the classifier.

Evaluating this classifier on the test set, we get an accuracy of 53.5%, somewhat disappointing. The main culprit was schools, which had an accuracy of only 22%, worse than random guessing. This may be due to the high variance in the energy usage for schools (as evidenced by the large standard deviations of the emission distributions for each state); it is possible that some schools look like restaurants and others look like dentists.

10.3 Classification: Multiclass Results

In this final phase, we discover and include two improvements to our classification pipeline.

The first is 10-fold cross validation, used to estimate error rates. The advantage of this method is that all observations are used for both training and validation, and each observation is used for validation exactly once [5]. Since k folds enables us to train on 90% of the data, we extend k-folds to the multi-class classification problem.

Second, we standardize our data (features have mean 0 and standard deviation 1). In our binary classification results, we noted that HMM features hurt accuracy on the normalized data (82% for just_usage vs 75% for usage_prob and 76% for usage_label). Intuitively, this shouldn't happen; they should just be ignored if not useful. After investigation, we realized that the HMM features (and usage_sum) were larger than the usage values, and therefore had higher variance, and so were implicitly weighted more heavily. We introduce standardization to address this issue. Figure 9 shows our results: HMM features no longer hurt accuracy: we achieve 88% for usage_prob with standardized features, vs. 75% with unstandardized features. Similar improvements hold for usage_label (88% vs 76%) and usage_prob_label (88% vs 75%). **Overall, the standardized results are much more stable, and accuracies are were at least as good (often much better) across feature sets**.

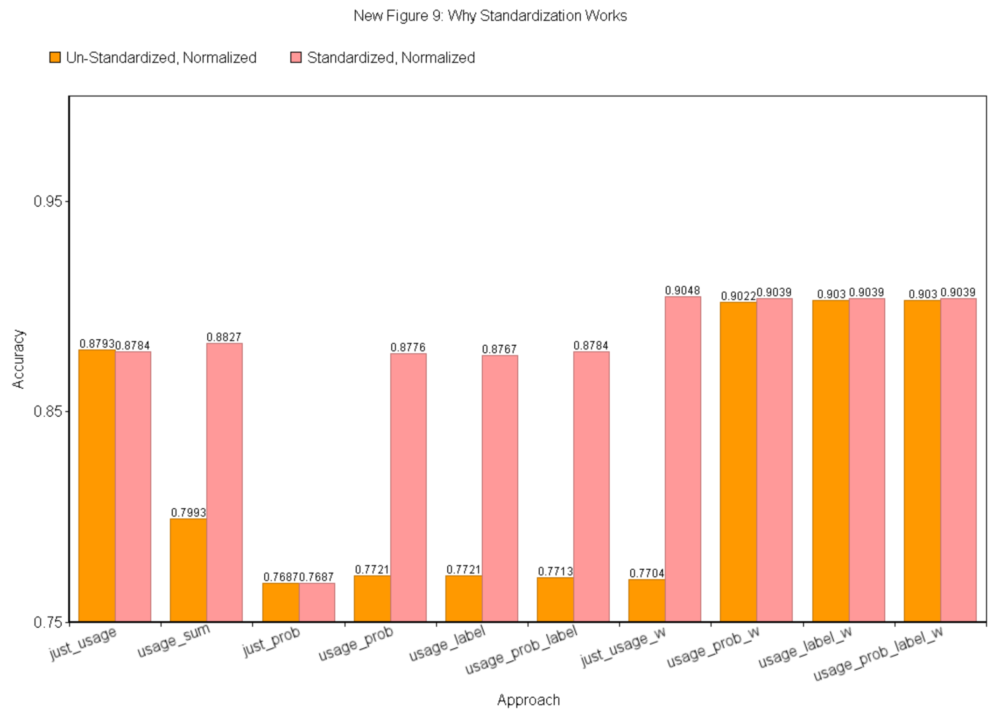

Figure 9: Why Standardization Works

Figure 10 shows the results of our logistic regression classifiers on both the binary and the multi-class problem for January 2014, with unnormalized and standardized data. Predictably, our error rate is higher in all cases on the multi-class problem. But the predictive power of the HMM probabilities is the least affected. As in the binary case, the weekend features were most informative for the multi-class problem. However, unlike the binary case, adding the HMM probabilities as features led to a noticeable increase in accuracy in the multi-class case. This suggests that when we try to classify a business as one of many different types, modeling energy usage as a HMM is indeed helpful.

11 Conclusion

In this project, we aimed to discover meaningful latent structure in energy usage data and solve an important classification problem along the way. To tackle the first problem, we built HMMs to learn hidden states and discover what look like open/close times–a result we can leverage in future work. To predict business types, we explore different feature sets and transformations to improve a logistic regression baseline, and we get good results. Furthermore, we try combining both techniques by plugging our HMM output into our classification pipeline. Although the binary classification results are only boosted minimally, the effect is stronger for the multi-class case, an encouraging result.

We did discover meaningful structure: dentists use more energy between 7 am and 5 pm, restaurants guzzle electricity during dinner time. On its own, the Bayesian classifier built from those generative models does not perform as well as LR, but the structure is useful in its own right, as discussed previously. To improve classification accuracy, we explore ten feature sets, two datasets, three transformations (normalization, standardization, k-folds validation), and one flavor of disaggregation. The features derived from the fitted HMMs slightly improve accuracy. Our weekend disaggregation improves accuracies across feature sets, as did unnormalized data, and feature standardization.

We also considered seasonal variation, repeating our analysis for data from July. In our latent structure mining, we found that generally higher usage levels in July, but qualitatively very similar hidden states.

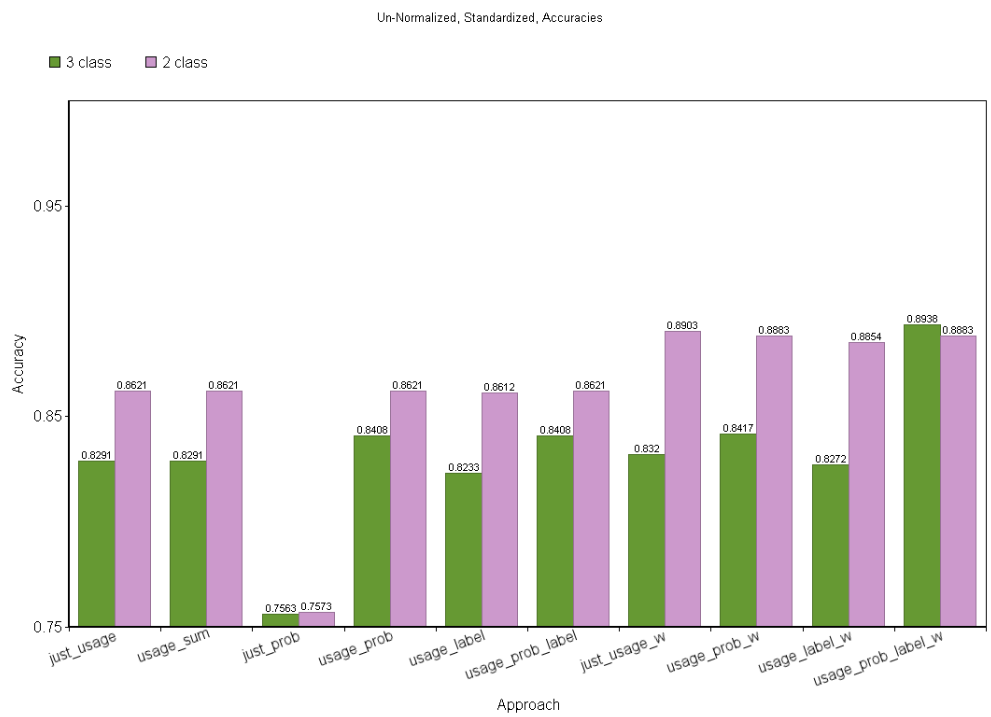

Figure 10: Performances of binary and multi-class LR Classifiers

In our classification effort, we found that our pipeline achieved better accuracy across the board on the July data (as did the HMM classifier alone).

Finally, we turned the difficulty to 11 by considering the 3-class problem of distinguishing between restaurants, dentists, and schools. Again, we learned meaningful states (and saw that schools burn lots of energy). This time though, our classification results were surprising. First, the HMM classifier had a much lower accuracy than logistic regression in the multiclass case, a reasonable failure given a more challenging problem. But we also discovered that the HMM features now led to a more substantial increase in accuracy of about 2%–an exciting and unexpected result, given the mediocre results we saw from combining these approaches in the binary case.

12 Future Work

As shown above, modeling energy usage as a HMM led us to find interesting patterns in the way that different types of businesses use energy throughout the day. For instance, restaurants used a lot of energy at night, while dentists and most schools did not. Utilities can take advantage of this information in a number of ways. They can effectively distribute energy among grids, which is especially important at peak times; in the evening, more energy should go to shopping districts and schools and less to business districts. Similarly, they could design district-specific rates to reduce peak loads.

We could extend this type of analysis further. One avenue for future work would be to for each business type, cluster businesses using their energy usage time series. This approach would discover more fine-grained patterns than before. For example, it may find that elementary schools use less energy than secondary schools in the evening, because there are fewer extracurricular activities.

References

[1] Christoph Flath, David Nicolay, Tobias Conte, Clemens van Dinther, and Lilia Filipova-Neumann. Cluster Analysis of Smart Metering Data, Business and Information Systems Engineering, Volume 4, Issue 1, February 2012, Pages 31-39, ISSN 1867-0202.

[2] Bing Dong, Cheng Cao, Siew Eang Lee, Applying support vector machines to predict building energy consumption in tropical region, Energy and Buildings, Volume 37, Issue 5, May 2005, Pages 545-553, ISSN 0378-7788.

[3] Geoffrey K.F. Tso, Kelvin K.W. Yau, Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks, Energy, Volume 32, Issue 9, September 2007, Pages 1761-1768, ISSN 0360-5442.

[4] Lawrence R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of the IEEE, Volume 77, No. 2, Pages 257-286, February 1989.

[5] Cross Validation, Wikipedia. https://en.wikipedia.org/wiki/Cross-validation