Building energy baseline model: next stage – non-linear methods with rich feature set

January 26, 2017

Contents

1 Problem specification

- using a set of features to predict hourly energy
- using a set of features to predict monthly energy we know the duration of the energy record
- using a set of features to predict hourly energy we do not know the duration of the energy record

2 Broad idea

From the discussing with Professor Matt Gormley, the broad approach should be: first create a rich feature set with all potentially related features included, and use a non-linear model on the rich feature set so that the training data can be nearly perfectly predicted. Then applying some regularization to also drive down the test error. Finally try to interpret the model by evaluating the accuracy drop by leaving each feature out, or by incrementally adding a feature in random order and evaluate the accuracy gain by adding that feature.

3 Collect input variables and representation

A list of variables that might be important are

- Environmental variable
  - Temperature (measured, average, or categorical):
    - outdoor air temperature
      - as numerical: mean [3, 6–8], degree-day [4, 10, 13], Radio Basis Function Kernel (RBFs) [15], exact [2, 9, 17]
      - as categorical variables [16]
    - indoor air temperature [7]
  - Humidity
    - relative humidity (RH) [3]
    - dew point temperature [2]
    - exponential smoothing applied to humidity with time constant of 24h [1]
  - Solar:

    * solar radiation (W/m^2) [3, 7]
    * solar flux [9]
    * solar aperture (m^2) [7], different in different time of year
    * solar gains ($Q_S = SI$, unit: W) [7]
  - Wind
    * speed [9]
    * velocity [1]

- Occupancy
  - Number of occupants [16]
  - Operation schedules [13]
  - Occupancy ratio (ratio of occupied vs non-occupied days) [12]
- Industry type
- Building construction
  - Detached vs apartment, categorical [16]
  - Construction material: wooden vs non-wooden [16]
- Time
  - day type (every-day, weekday, weedend) [6]
  - hour of day ([6, 15], [5] mean-week and day-time-temperature regression model)
  - day of week ([5] mean-week, day-time-temperature, and LBNL regression model)
  - time lag (k), the number of previous readings to include in the model [7]
  - unit circle representation of time of day, week, month, and year [1]
- Energy
  - power (W, it's an auto-regressive component: use energy to predict energy) [7]([11] has some experiment about prediction of different time horizon using different time resolution)
  - fuel type: Electric vs non-electric [16]
- Floor area [16]

- Building dynamics
  - Heat loss coefficient (W/m^2K) [16]
  - Equivalent leakage area (cm^2/m^2) [16]
- Retrofit type / time
  - pre-retrofit period [8]

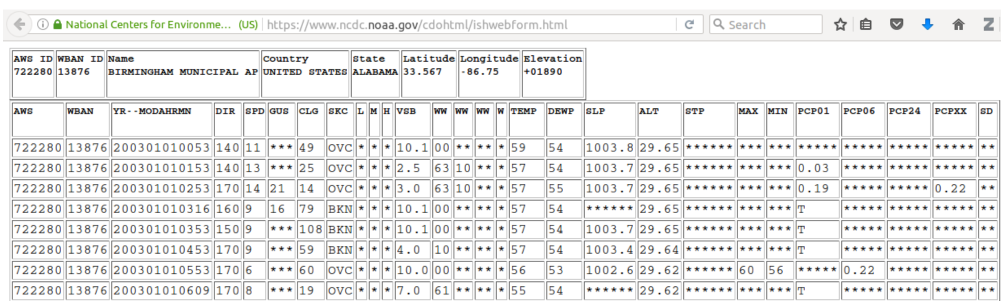
4 Data sources

4.1 Environment data sources

In the previous stage of the work, the data source for temperature, the only environmental variable is retrieved from the pisystem, whose source is weather underground web interface.

4.1.1 Integrated Surface Global Hourly Data from NOAA

- sample



AWS ID	WBAN ID	Name	Country	State	Latitude	Longitude	Elevation
722280	13876	BIRMINGHAM MUNICIPAL AP	UNITED STATES	ALABAMA	33.567	-86.75	+01890

AWS	WBAN	YR--MODAHRMN	DIR	SPD	GUS	CLG	SKC	L	M	H	VSB	WW	WW	WW	W	TEMP	DEWP	SLP	ALT	STP	MAX	MIN	PCP01	PCP06	PCP24	PCPXX	SD
722280	13876	200301010053	140	11	***	49	OVC	*	*	*	10.1	00	**	**	*	59	54	1003.8	29.65	******	***	***	*****	*****	*****	*****	**
722280	13876	200301010153	140	13	***	25	OVC	*	*	*	2.5	63	10	**	*	57	54	1003.7	29.65	******	***	***	0.03	*****	*****	*****	**
722280	13876	200301010253	170	14	21	14	OVC	*	*	*	3.0	63	10	**	*	57	55	1003.7	29.65	******	***	***	0.19	*****	*****	0.22	**
722280	13876	200301010316	160	9	16	79	BKN	*	*	*	10.1	00	**	**	*	57	54	******	29.65	******	***	***	T	*****	*****	*****	**
722280	13876	200301010353	150	9	***	108	BKN	*	*	*	10.1	00	**	**	*	57	54	1003.7	29.65	******	***	***	T	*****	*****	*****	**
722280	13876	200301010453	170	9	***	59	BKN	*	*	*	4.0	10	**	**	*	57	54	1003.4	29.64	******	***	***	T	*****	*****	*****	**
722280	13876	200301010553	170	6	***	60	OVC	*	*	*	10.0	00	**	**	*	56	53	1002.6	29.62	******	60	56	*****	0.22	*****	*****	**
722280	13876	200301010609	170	8	***	19	OVC	*	*	*	7.0	61	**	**	*	55	54	******	29.62	******	***	***	T	*****	*****	*****	**

Figure 1: NOAA Integrated Surface Global Hourly Data

- download: `ftp://ftp.ncdc.noaa.gov/pub/data/noaa/`
- Data are ordered by year and station, each data file contains weather station identifier (USAF, and WBAN)
- relevant fields: wind direction and speed, sky cover condition (clear, overcast, scattered, etc.), temperature, dew point, precipitation.
- time resolution: 1 to 2 observations per hour

4.2 LDAS

5 Non-linear models

- Neuron Network with 1-2 hidden layer
- Support Vector Regression with RBF kernel
- Random forest regression [14]
- piecewise linear regression as baseline (it's a simple non-linear model, but not expressive enough)

References

[1] Matthew Brown, Chris Barrington-Leigh, and Zosia Brown. Kernel regression for real-time building energy analysis. *Journal of Building Performance Simulation*, 5(4):263–276, 2012.

[2] Li-Juan Cao and Francis Eng Hock Tay. Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on neural networks*, 14(6):1506–1518, 2003.

[3] Bing Dong, Cheng Cao, and Siew Eang Lee. Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings*, 37(5):545–553, 2005.

[4] Margaret F Fels. Prism: an introduction. *Energy and Buildings*, 9(1-2):5–18, 1986.

[5] Jessica Granderson. Evaluation of the predictive accuracy of five whole building baseline models, 2014.

[6] JS Haberl and S Thamilseran. A bin method for calculating energy conservation retrofit savings in commercial buildings, 1994.

[7] Stig Hammarsten. A critical appraisal of energy-signature models. *Applied Energy*, 26(2):97–110, 1987.

[8] John Kelly Kissock. *A methodology to measure retrofit energy savings in commercial buildings*. PhD thesis, UMI, 2008.

[9] David JC MacKay. Bayesian non-linear modeling for the prediction competition. In *Maximum Entropy and Bayesian Methods*, pages 221–234. Springer, 1996.

[10] Energy Star Portfolio Manager. Climate and weather. `https://portfoliomanager.energystar.gov/pdf/reference/Climate%20and%20Weather.pdf`. Accessed: 2016-10-13.

[11] Elena Mocanu, Phuong H Nguyen, Madeleine Gibescu, and Wil L Kling. Deep learning for estimating building energy consumption. *Sustainable Energy, Grids and Networks*, 6:91–99, 2016.

[12] Ari Rabl and Anne Rialhe. Energy signature models for commercial buildings: test with measured data and interpretation. *Energy and buildings*, 19(2):143–154, 1992.

[13] T Agami Reddy, Namir F Saman, David E Claridge, Jeff S Haberl, WD Turner, and AT Chalifoux. Baselining methodology for facility-level monthly energy use-part 1: Theoretical aspects. *TRANSACTIONS-AMERICAN SOCIETY OF HEATING REFRIGERATING AND AIR CONDITIONING ENGINEERS*, 103:336–347, 1997.

[14] Wikipedia. Random forest. `https://en.wikipedia.org/wiki/Random_forest`. Accessed: 2016-12-17.

[15] Matt Wytock and J Zico Kolter. Contextually supervised source separation with application to energy disaggregation. *arXiv preprint arXiv:1312.5023*, 2013.

[16] Zhun Yu, Fariborz Haghighat, Benjamin C.M. Fung, and Hiroshi Yoshino. A decision tree method for building energy demand modeling. *Energy and Buildings*, 42(10):1637 – 1646, 2010.

[17] Yuna Zhang, Zheng O'Neill, Bing Dong, and Godfried Augenbroe. Comparisons of inverse modeling approaches for predicting building energy performance. *Building and Environment*, 86:177 – 190, 2015.