

ELECTRICITY DATA ANALYSIS BASED ON SPARSE CODING

Lujun Zhao

Join with Li Zhou

Supervised by: Professor Weiguo Gao

May 11, 2015

School of Mathematical Sciences

Fudan University

INTRODUCTION

- Energy consumption of both residential and commercial buildings has steadily increased, reaching figures between 20% and 40% in developed countries.
- The rapidly growing has already raised concerns over supply difficulties, exhaustion of energy resources and heavy environmental impacts.
- There is an urgent need to develop technologies that examine energy usage in homes.

SOME STATISTICAL DATA

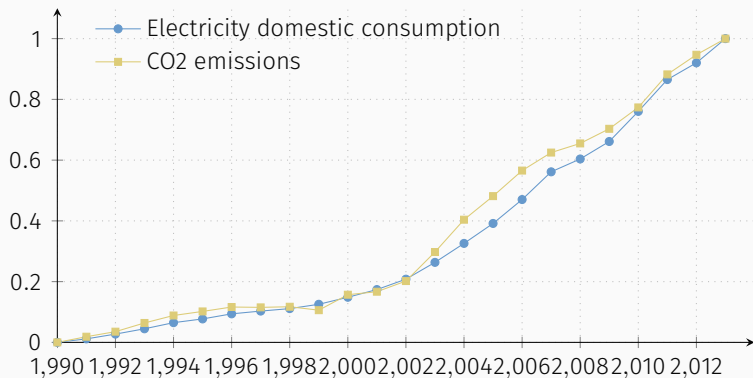


Figure 1: Electricity domestic consumption and CO2 emissions in China. **Source:** Enerdata.

- Find behavioral patterns and outlier detection.
- Helpful for building energy conservation services.
- Give residents feedback on energy consumption.
- Encourage energy efficient behaviors.

Classification and clustering algorithms:

- K-Means [Chen and Cook]
- CART [LIU, CHEN, MORI and KIDA]
- DBSCAN [Khan, Capozzoli, Corgnati and Cerquitelli]

Statistical Approach:

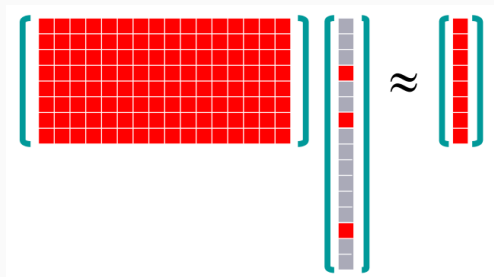
- Generalized Extreme Studentized Deviate (GESD) [John E. Seem]
- boxplot statistical method [John E. Seem]

OUR MODEL

In this report, we present a novel approach for finding outliers and pattern recognition based on **sparse model** that include the follow two steps:

- Sparse Coding
- Dictionary Learning

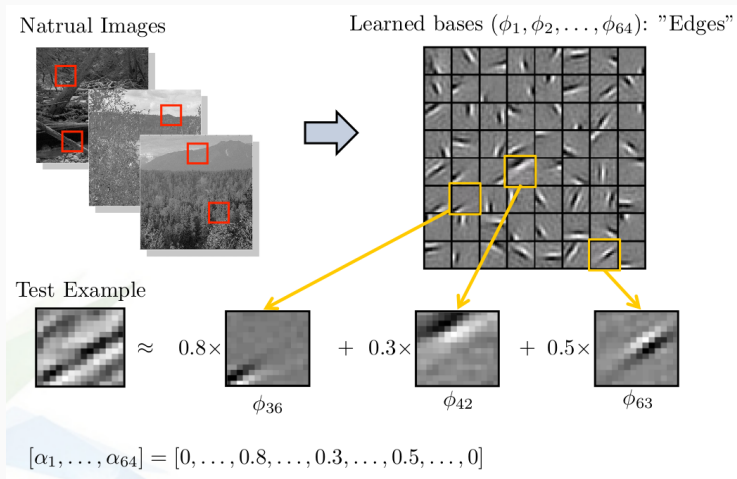
Applications that use this model are many and include compression, feature extraction, image denoising and more.



$$D \quad x \quad \approx \quad y$$

x is sparse. **Sparsity** implies many zeros in a vector.

SPARSE CODING OF IMAGES



Generally, we expect sparse solutions, a obvious idea is find the solution with the fewest number of non-zero coefficients:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{y} = \mathbf{D}\mathbf{x}, \quad (1)$$

or

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 < \epsilon \quad (2)$$

Here $\|\cdot\|_0$ is the l^0 norm, counting the number of non-zeros of a vector.

Unfortunately, (1) is not a convex optimisation problem. In fact, it is an NP-hard problem in general [Davis, Mallat and Avellaneda].

Geoff Davis, Stephane Mallat, and Marco Avellaneda. Adaptive greedy approximations. Constructive approximation, 13(1):57–98, 1997.

There have many algorithms to find approximating solutions.

Greedy algorithms:

- *Matching Pursuit* (MP)[Mallat and Zhang]
- *Orthogonal Matching Pursuit* (OMP)[Pati, Rezaiifar and Krishnaprasad]
- ...

St éphane G Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397–3415, 1993.

Yagyensh Chandra Pati, Ramin Rezaiifar, and PS Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers*, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on, pages 40–44. IEEE, 1993.

Another simple, yet surprisingly effective, way to do so is l^1 minimisation or *Basis Pursuit*[Chen, Donoho, and Saunders].

Basis Pursuit:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{D}\mathbf{x} \quad (3)$$

This is a convex optimisation problem and can be transformed to a linear programming problem that can be solved fairly by simplex method or interior point method.

Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1998.

If the data has some noises, we should apply *Basis Pursuit Denoising* (BPDN):

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 < \epsilon \quad (4)$$

or equivalent

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \quad (5)$$

(5) is also known as *Lasso*[Tibshirani] in statistics.

Robert Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), pages 267–288, 1996.

WHY DOES l^1 -NORM INDUCE SPARSITY?

The geometric explanation:

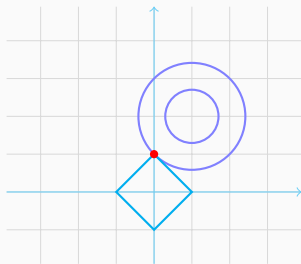


Figure 2: l^1 -norm

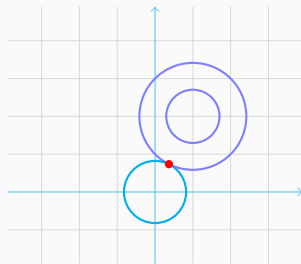


Figure 3: l^2 -norm

Rigorous proof by Candès, Romberg and Tao.

Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. Information Theory, IEEE Transactions on, 52(2):489–509, 2006.

How can we get the dictionary **D**?

1. Predefined dictionary

- DCT Basis
- Wavelet
- ...

2. Adaptive dictionary - learn from data

- K-SVD (l^0 -norm)
- Online dictionary learning (l^1 -norm)
- ...

PREDEFINED DICTIONARY

- DCT Basis: used in JPEG image compression.
- Wavelet: used in JPEG-2000.



K-SVD is a generalization of K-means algorithm and also an iterative method that alternates between sparse coding and The update of the dictionary[Aharon, Elad and Bruckstein].

Objective function:

$$\min_{\mathbf{D}, \mathbf{X}} \{ \|\mathbf{Y} - \mathbf{DX}\|_F^2 \} \quad \forall i, \|\mathbf{x}_i\|_0 \leq T_0. \quad (6)$$

Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. Signal Processing, IEEE Transactions on, 54(11):4311–4322, 2006.

K-MEANS VS K-SVD

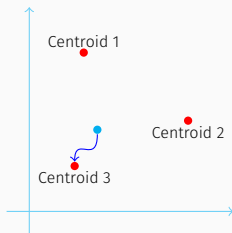


Figure 4: K-Means

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

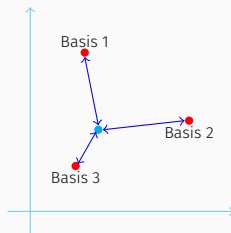


Figure 5: K-SVD

$$\begin{bmatrix} 0.3 \\ 0.1 \\ 0.6 \end{bmatrix}$$

We will use the online dictionary learning algorithm[Mairal, Bach, Ponce and Sapiro] in our model to analyze energy data. This algorithm is also widely used in many libraries such as scikit-learn, SPAMS, etc.

Different from K-SVD, the online dictionary learning algorithm uses l^1 rather than l^0 penalties.

Objective function:

$$\min_{\mathbf{D} \in \mathcal{C}, \alpha \in \mathbb{R}^{k \times n}} \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{2} \|\mathbf{x}_i - \mathbf{D}\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1 \right). \quad (7)$$

Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. In Proceedings of the 26th Annual International Conference on Machine Learning, pages 689–696. ACM, 2009.

Similar with K-SVD, the Online Dictionary Learning is also an iterative method.

In Sparse Coding, draws one element \mathbf{x}_t at a time, then compute using LARS:

$$\alpha_t \triangleq \operatorname{argmin}_{\alpha \in \mathbb{R}^k} \frac{1}{2} \|\mathbf{x}_t - \mathbf{D}_{t-1}\alpha\|_2^2 + \lambda \|\alpha\|_1. \quad (8)$$

In Dictionary Learning, compute \mathbf{D}_t using gradient descent:

$$\begin{aligned} \mathbf{D}_t &\triangleq \operatorname{argmin}_{\mathbf{D} \in \mathcal{C}} \frac{1}{t} \sum_{i=1}^t \frac{1}{2} \|\mathbf{x}_i - \mathbf{D}\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1, \\ &= \operatorname{argmin}_{\mathbf{D} \in \mathcal{C}} \frac{1}{t} \left(\frac{1}{2} \operatorname{Tr}(\mathbf{D}^T \mathbf{D} \mathbf{A}_t) - \operatorname{Tr}(\mathbf{D}^T \mathbf{B}_t) \right). \end{aligned} \quad (9)$$

EXPERIMENT

Through some city electric power company, we collect the data of residents electricity consumption range from 2013-1-1 to 2014-6-23 , a total of 538 days.

Table 1: Consumption data

User	Day 1	Day 2	Day 3	Day 4	Day 5	...	Day 535	Day 536	Day 537	Day 538
1	11.37	21.03	12.22	14.79	28.59	...	13.41	10.11	6.71	8.06
2	38.14	84.02	80.92	37.64	2.50	...	5.24	8.54	6.34	6.44
3	15.25	28.51	16.20	18.63	23.72	...	7.10	6.15	6.38	7.51
4	1.80	2.70	19.31	48.26	2.22	...	24.37	17.01	16.39	17.25
5	2.15	11.03	2.63	75.19	87.84	...	15.94	17.15	16.53	13.02
...
4314	31.92	22.99	20.86	11.10	17.49	...	12.30	11.60	9.31	8.82
4315	8.48	5.61	5.85	6.19	4.92	...	5.61	4.75	5.57	9.21
4316	2.33	2.15	1.07	1.45	1.43	...	2.07	1.98	3.15	2.90

We transform the formal data of each row to a pattern organized by week and the new row of data is made up of 7 days.

Table 2: Weekly consumption data

User	Week	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
1	1	11.37	21.03	12.22	14.79	28.59	16.64	7.80
	2	6.63	23.18	9.17	14.36	21.91	5.11	7.79
	3	5.64	11.45	7.48	10.13	7.27	7.11	7.38
	4	4.96	6.14	9.47	11.08	10.33	5.70	9.94
	5	4.08	4.95	3.97	8.94	6.91	6.26	4.90

...
4316
	72	1.32	1.89	1.91	2.02	1.20	2.97	2.17
	73	2.52	1.39	3.58	2.56	3.52	3.79	6.09
	74	3.54	5.60	5.21	3.66	3.06	3.68	1.92
	75	2.79	2.07	2.14	3.65	2.39	2.18	1.91
	76	6.44	2.58	6.76	2.77	3.01	1.73	1.93

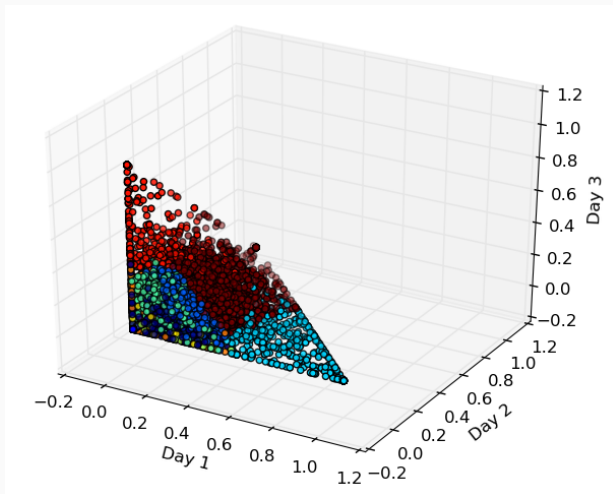


Figure 6: K-Means

Use Sparse Coding first, then apply K-Means to sparse matrix.

Test on digits dataset:

- K-Means: 467/1797
- This method: 280/1797

Test on iris dataset:

- K-Means: 16/150
- This method: 16/150

The each column of \mathbf{D} is a energy pattern. One resident's consumption data can be represented as a linear combination of some patterns. we consider an energy pattern d_j to be an outlier, if this energy pattern is rarely be used in the linear combination.

We use the score below to describe the pattern's degree of usage:

$$f(j) = \sum_{k=1}^N g\left(\frac{\sum_{i \in U_k} \frac{|\alpha_{ij}|}{\sum_{q=1}^M |\alpha_{iq}|}}{|U_k|}\right)$$

where U_k is the set of rows belong to user k , $|U_k|$ is the number of elements in the set U_k , N is the number of users, M is the number of basis and $g(x)$ is a function we choose to reduce single user affect in the outliers score. For each pattern d_j , it has a score $f(j)$.

THE BASIS MOST FREQUENTLY USED

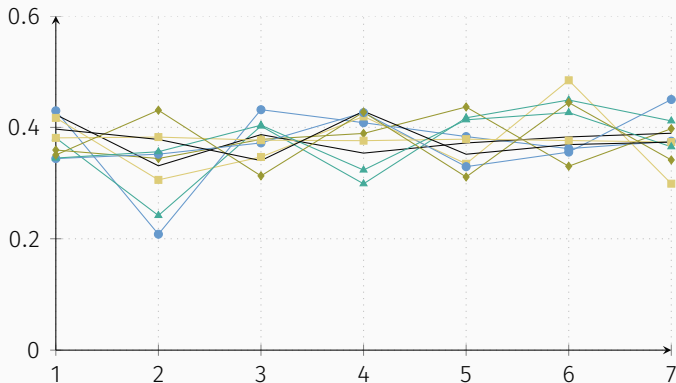


Figure 7: The basis most frequently used

EXAMPLE

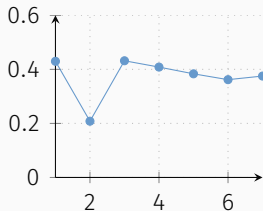


Figure 8: One of most used basis

User	Week	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
200	54	109.29	69.47	107.25	144.24	99.86	95.21	109.30
1537	59	107.50	64.14	103.02	95.95	91.43	106.42	85.58
1675	20	147.68	121.69	156.19	148.74	125.92	127.63	148.35
1769	64	80.19	46.58	81.96	86.65	72.62	82.17	80.93
1474	60	124.23	63.35	128.49	119.88	127.36	104.49	114.25

Table 3: The users use this basis

We can easily find from the two figures that they have the same behavior as the pattern, so we can use the pattern to represent these users' behavior in this week.

ANOTHER EXAMPLE

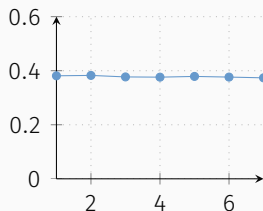


Figure 9: Another basis most used

User	Week	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
739	29	190.28	193.52	182.01	189.68	200.06	194.81	193.41
	66	140.33	141.95	140.46	141.91	146.73	140.49	139.42
	39	144.40	145.02	146.36	145.97	148.70	151.53	145.64
1469	32	243.09	239.92	234.66	212.71	229.12	228.68	221.51
	29	239.88	240.09	224.31	236.91	230.27	220.61	218.10

Table 4: The users use this basis

We see that these users are nearly unchangeable in one week and have very large consumption. If you observe these users' data in other weeks, for example, users 1469, you can find that the data is also similar with this pattern.

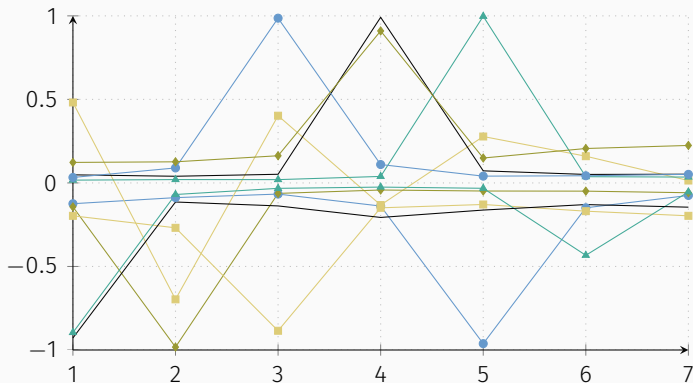


Figure 10: The basis of outliers

EXAMPLE

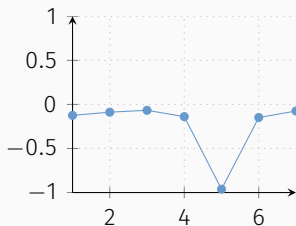


Figure 11: One basis of outliers

User	Week	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
1768	57	35.44	21.80	11.23	14.93	51.19	33.71	31.92
1810	33	16.73	9.30	12.57	8.92	31.95	9.42	3.82
19	33	30.90	27.32	10.18	16.00	59.64	26.22	14.84
4	55	32.59	39.48	1.59	33.45	69.56	40.48	32.84
25	6	112.36	0.00	1.94	39.05	84.06	44.72	13.00

Table 5: The users use this basis

CONCLUSION

- In this report, we presented a novel approach for analyzing energy consumption data.
- By using Sparse Coding method, we detect the outliers which less used.
- This will be helpful for building energy conservation services and give residents feedback about energy efficient behaviors.

- Improve model to adapt the energy data.
- Test using more data in more powerful machine or hadoop cluster.

Thanks to:

- My advisor: Prof. Weiguo Gao
- My partner: Li Zhou

THANK YOU!