Power Curve Modeling

Haiyang Zheng

Intelligent Systems Laboratory
The University of Iowa

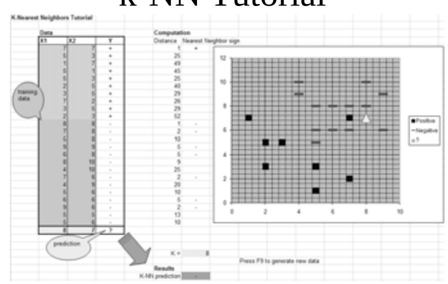
Outline

- k-NN (k-nearest neighbor) algorithm
- Wind farm power modeling based on k-NN and PCA (Principle Component Analysis)
- Parametric models of wind farm power

k-NN algorithm

- Represent each instance in a multi-dimension space.
- Divide the entire data set into training and test data sets.
- Given a test instance, a distance metric is computed between the test instance and all training instance, then the k nearest neighbors are selected from the training data.
- Compute the average values of the k nearest neighbors. This value becomes the predicted value for the test instance.

k-NN Tutorial



Visit the complete version of this tutorial in http://people.revoledu.com/kardi/tutorial/KNN/index.html

Small Example of k-NN

Assume x1 and x2 are the two measurements of the final product quality, how to use k-NN for product quality judgment?

No	X1= Strength	X2= Smooth	Y=Quality
1	5	5	Good
2	1	2	Bad
3	6	7	Good
4	3	5	Bad
5	2	3	?

Algorithm Schemes

- Different distance metrics are used, e.g. Euclidean, Manhattan.
- The parameter k is significant in k-NN algorithm and its best value depends on the data structure and conditions.
- Euclidean distance metric is selected and k is set to 100 based on the model’s prediction accuracy.

Euclidean distance

- Two data point P and Q represented in Euclidean n-space,

$$P = (p_1, p_2, ..., p_n) \qquad Q = (q_1, q_2, ..., q_n)$$

The Euclidean distance between P and Q is express as following:

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + ... + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$$

The University of Iowa   Intelligent Systems Laboratory

K=1, 1-nearest neighbor

- Data set 1-4 are the training data set, and data set 5 is the test data set.
- Compute the distance between test data and training data.

$$D_{15} = \sqrt{(5-2)^2 + (5-3)^2} = 3.6 \qquad D_{25} = \sqrt{(1-2)^2 + (2-3)^2} = 1.4$$

$$D_{35} = \sqrt{(6-2)^2 + (7-3)^2} = 5.6 \qquad D_{45} = \sqrt{(3-2)^2 + (5-3)^2} = 2.2$$

- D_{25} is the smallest, and thus data point 2 is chosen as the 1-nearest neighbor.
- Thus the quality is bad.

The University of Iowa   Intelligent Systems Laboratory

K=3, 3-nearest neighbors

$$D_{15} = \sqrt{(5-2)^2 + (5-3)^2} = 3.6 \qquad D_{25} = \sqrt{(1-2)^2 + (2-3)^2} = 1.4$$

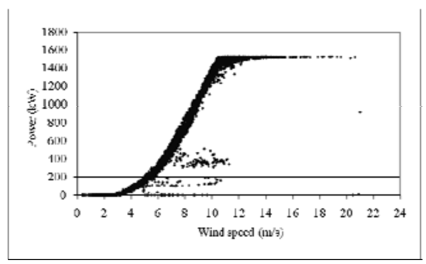
$$D_{35} = \sqrt{(6-2)^2 + (7-3)^2} = 5.6 \qquad D_{45} = \sqrt{(3-2)^2 + (5-3)^2} = 2.2$$

- Data point 1, 2 and 4 are chosen as 3-nearest neighbors. Two training data vote for bad, and one for good. The quality of 5 is still bad.
- Assume the value of good is 1, and bad is 0.

$$y_5 = (1 + 1 + 0)/3 = 0.6$$

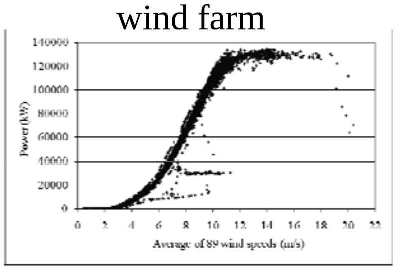
The University of Iowa   Intelligent Systems Laboratory

A typical power curve of a single turbine



The University of Iowa   Intelligent Systems Laboratory

Cumulative power curve of the wind farm



The wind farm contain 100 turbines, 89 turbines with good quality data are selected

The University of Iowa   Intelligent Systems Laboratory

Data Description

Data set	Start Time Stamp	End Time Stamp	Description
1	1/1/06 12:00 AM	1/31/06 11:50 PM	Total data set; 4347 observations
2	1/1/06 12:00 AM	1/25/06 6:20 PM	Training data set; 3476 observations
3	1/25/06 6:30 PM	1/31/06 11:50 PM	Test data set; 871 observations

10-minute SCADA data

The University of Iowa   Intelligent Systems Laboratory

Data Format

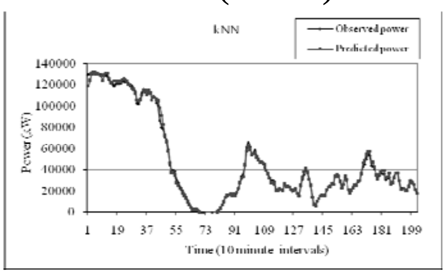
X1	X2	….	X89	Y
Wind speed of Turbine 1	Wind speed of Turbine 2	Wind speed of Turbine *i*	Wind speed of Turbine 3	Wind farm power

Preprocess the data and prepare them into the format for Data Mining Software, e.g. Weka and Statistica

Data in Excel

TimeStamp	turbine78	turbine79	turbine80	turbine81	turbine82	turbine83	turbine84	turbine85	turbine86	turbine87	turbine88	turbine89	Total Power
1/1/06 12:00 AM	7.63	8.42	8.17	8.52	8.17	7.31	7.84	7.82	7.02	9.61	8.51	8.87	68636.81
1/1/06 12:10 AM	7.69	8.35	8.66	9.04	8.46	6.93	7.82	7.86	7.99	9.39	8.81	8.81	70714.8
1/1/06 12:20 AM	7.9	8.35	8.91	8.95	8.49	7.13	7.92	7.84	8.26	9.5	8.75	8.47	73608.74
1/1/06 12:30 AM	8.26	8.38	9.12	9.11	8.36	7.39	8.25	7.84	8.07	9.14	8.62	8.41	73481.25
1/1/06 12:40 AM	8.1	8.3	9.26	9.01	8.08	7.58	8.14	7.82	8.32	9	8.43	8.43	76098.88
1/1/06 12:50 AM	7.58	8.3	9.07	8.9	8.25	7.29	8.02	7.98	8.24	8.67	8.33	8.36	72087.7
1/1/06 1:00 AM	7.34	8.08	8.82	8.41	8.09	7.46	7.83	7.63	7.8	8.85	8.29	8.39	71699.46
1/1/06 1:10 AM	7.93	8.37	8.92	8.74	7.83	7.12	7.92	7.97	7.66	9.23	8.67	8.91	74591.62
1/1/06 1:20 AM	7.95	8.21	8.89	8.33	8.07	7.01	8.03	7.63	8.01	9.32	8.68	8.54	74218.96
1/1/06 1:30 AM	8.03	8.25	8.82	8.68	8.06	6.99	8.17	7.5	7.77	9.51	8.93	8.62	78886.46
1/1/06 1:40 AM	7.89	8.16	8.73	8.36	7.97	6.84	8.02	7.69	7.55	9.5	8.65	8.52	74139.21
1/1/06 1:50 AM	8.29	8.15	8.75	8.28	8.07	6.81	7.98	7.74	7.76	9.41	8.61	8.8	73640.2
1/1/06 2:00 AM	8.43	8.21	8.63	8.32	7.9	7.06	8.18	8.05	7.47	9.48	8.93	8.9	76455.75
1/1/06 2:10 AM	8.24	8.39	8.63	8.47	8.04	7.09	8.18	8.01	7.77	9.7	8.81	9.54	79084.05
1/1/06 2:20 AM	8.53	8.61	8.77	8.68	8.02	7.36	8.32	8.26	7.86	9.91	9.08	9.61	80558.84
1/1/06 2:30 AM	8.51	8.8	8.72	8.54	8.42	7.13	8.54	8.3	7.76	10.02	9.13	9.62	81226.97
1/1/06 2:40 AM	7.54	8.19	8.63	7.98	8.56	6.49	8.31	8.11	7.58	9.75	8.93	8.81	73501.89
1/1/06 2:50 AM	7.24	8.47	8.96	8.43	8.22	6.35	8	8.05	7.22	9.16	8.73	8.55	69402.39
1/1/06 3:00 AM	7	8.63	8.58	8.47	7.53	6.36	7.26	7.53	6.95	8.72	8.48	8.32	62922.83
1/1/06 3:10 AM	7.03	7.94	8.24	7.9	7.3	6.44	7.07	7.63	7.4	8.83	8.01	7.98	57995.75
1/1/06 3:20 AM	7.23	8.1	8.7	8.38	6.8	6.36	7.06	7.46	7.71	8.43	7.86	8.37	57729.55
1/1/06 3:30 AM	6.9	7.99	8.7	7.95	6.65	6.19	6.64	7.4	7.7	8.41	7.77	7.73	55198.77
1/1/06 3:40 AM	7.27	8.03	8.67	7.83	6.67	6.07	6.69	7.38	7.54	8.61	7.61	7.81	53366.44
1/1/06 3:50 AM	7.07	7.6	8.45	8.08	6.58	6.53	6.7	7.23	6.89	8.31	7.36	7.62	51685.44
1/1/06 4:00 AM	6.58	7.47	7.87	7.94	6.49	6.24	6.32	7.21	6.86	8.19	7.35	7.58	49744.53
1/1/06 4:10 AM	6.55	7.19	7.69	7.95	6.63	5.8	6.42	7.38	6.85	8.04	7.44	7.7	48688.23
1/1/06 4:20 AM	6.11	6.92	7.47	7.65	6.79	5.27	6.34	7.1	7.1	7.66	7.45	7.48	45344.89

k-NN (k=100)



MAE (Means absolute error): 2872 kW
Std (Standard deviation of absolute error): 2949 kW

Small Example of Wind farm Power Modeling

Time Stamp	Turbine_1	Turbine_2	Turbine_3	Turbine_4	Total Power
1/1/06 12:00 AM	7.96	8.92	8.78	7.17	3556.85
1/1/06 12:10 AM	8.35	8.49	9	6.86	3514.91
1/1/06 12:20 AM	8.5	8.4	9.06	6.89	3621.85
1/1/06 12:30 AM	8.34	8.4	9.12	7.02	3499.33
1/1/06 12:40 AM	7.98	8.5	9.44	6.75	?

K=2, Euclidean distance

Euclidean distance	1/1/06 12:00 AM	1/1/06 12:10 AM	1/1/06 12:20 AM	1/1/06 12:30 AM
1/1/06 12:40 AM	0.89	0.58	0.67	0.56

The data at time 1/1/06 12:10 AM and 1/1/06 12:30 AM are chosen as 2-nearest neighbors. The average total power at these two time points is the prediction value of wind farm power.

$$Totalpower = (3514.91 + 3499.33)/2 = 3507.12\,kW$$

The actual total power at time 1/1/06 12:40 AM is 3512.05 kW.

Feature reduction and transformation

- The k-NN model built before has 89 inputs, and thus the input dimensionality need to be reduced.
- The principal component analysis (PCA) was chosen to do feature transform and reduction.
- The PCA expresses the variance-covariance structure of a set of variables by a few linear combinations

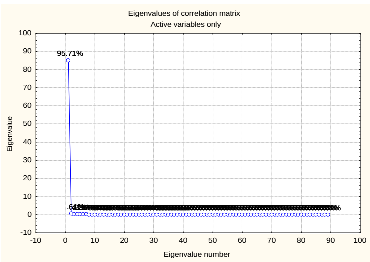
Basic steps of PCA

- Compute a correlation matrix.
- Compute the eigenvectors and eingenvalues of the correlation matrix.
- Select the components to form an eingenvector.
- Derive the new data comprised of the principal component of the original data.

Eigen values of the correlation matrix and the related statistics

Value Number	Eigen value	Total Variance (%)	Cumulative Eigen value	Cumulative (%)
1	85.17	95.70	85.17	95.70
2	0.54	0.61	85.72	96.31
3	0.36	0.41	86.09	96.73
4	0.20	0.23	86.29	96.96
5	0.18	0.20	86.48	97.17
6	0.15	0.17	86.63	97.34
7	0.18	0.15	86.77	97.49
8	0.13	0.14	86.90	97.64
9	0.11	0.12	87.01	97.77
10	0.09	0.11	87.11	97.88

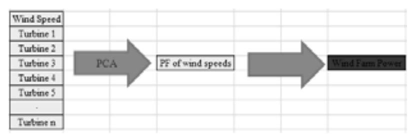
Plot of Eigenvalues of correlation matrix



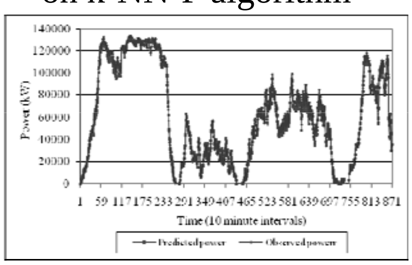
k-NN-P (integrated k-NN and PCA) Model

$$P_1 = \partial_1^1 x_1 + \partial_2^1 x_2 + \ldots + \partial_n^1 x_n$$
$$P_2 = \partial_1^2 x_1 + \partial_2^2 x_2 + \ldots + \partial_n^2 x_n$$
$$\ldots$$
$$P_n = \partial_1^n x_1 + \partial_2^n x_2 + \ldots + \partial_n^n x_n$$

P_i: the PF (principal factor)
x_i: the wind speed of 89 turbines
∂_i^i: the weight for feature transformation

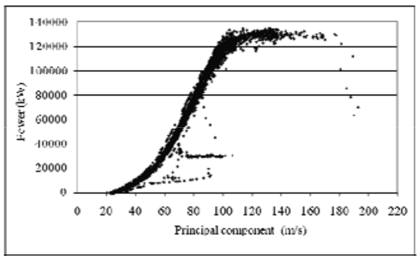


Wind farm power model based on k-NN-P algorithm



MAE: 2255 kW; Std: 2174 kW

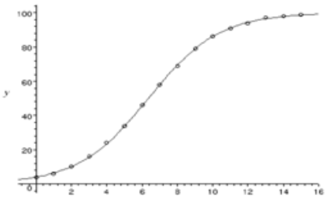
Power curve based on PCA



Observed power curve as the function of the principal component derived from 89 wind speeds.

Nonlinear parametric modeling of wind farm power curves

- Logistic function



$$y = f(x,\boldsymbol{\theta}) = a\frac{1+me^{-x/\tau}}{1+ne^{-x/\tau}}$$

$$\boldsymbol{\theta} = (a, m, n, \tau)$$

The University of Iowa — Intelligent Systems Laboratory

Learning Parametric model from training data

$$y = f(x,\boldsymbol{\theta}) = a\frac{1+me^{-x/\tau}}{1+ne^{-x/\tau}} \qquad \boldsymbol{\theta} = (a, m, n, \tau)$$

x: is the principal component of 89 wind speeds

y: the power of the wind farm

θ: is a 4-dimension vector parameter of logistic function that determines the shape of the power curve

$$S_{(x,y)} = \sum_{i=1}^{N}[a\frac{1+me^{-x(i)/\tau}}{1+ne^{-x(i)/\tau}} - y(i)]^2 \qquad \hat{\boldsymbol{\theta}} = \underset{a,m,n,\tau}{\operatorname{argmin}} S_{(x,y)}\left(x(1), y(1);...;x(N), y(N)|a,m,n,\tau\right)$$

cost function

The University of Iowa — Intelligent Systems Laboratory

The basic steps of the evolutionary strategy algorithm

1: Initialize μ individuals (candidate vector parameter) to form the initial parent population.

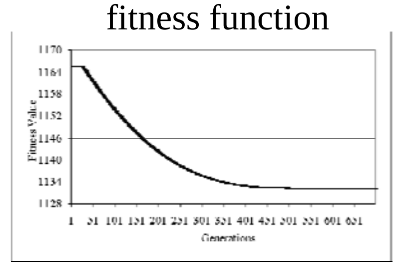
2: Repeat until the stopping criteria are satisfied.

- 2.1: Select from the parent population and recombine 2 parents λ times to generate λ children.
- 2.2: Mutate λ children.
- 2.3: Select the best μ individuals from the children and parent pool based on the fitness function values.
- 2.4: Use these μ selected individuals as parents for the next generation.

3: Apply one of the stopping criteria, e.g., the allowable number of generations.

The University of Iowa — Intelligent Systems Laboratory

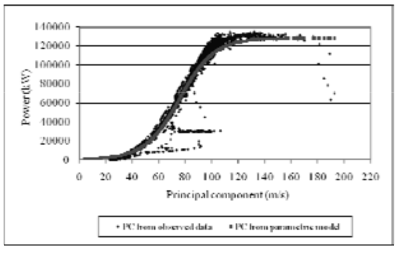
Convergence process of the fitness function



$$\hat{\boldsymbol{\theta}} = (128545.6123, 0.7106, 320.8248, 13.1239)$$

The University of Iowa — Intelligent Systems Laboratory

Power curve generated from the observed and the parametric model



The University of Iowa — Intelligent Systems Laboratory