Slide 1

EUNITE Competition: Electricity Load Forecasting

Ming-Wei Chang, Bo-Juen Chen and Chih-Jen Lin

Department of Computer Science and
Information Engineering
National Taiwan University

EUNITE 2001, Tenerife, Spain, December 14, 2001

Slide 2

Outline

- Support vector regression (SVR) for time series prediction
- Data preprocessing and some practical issues
- Implementations and experiments
- Results
- Discussion

Slide 3

The Problem

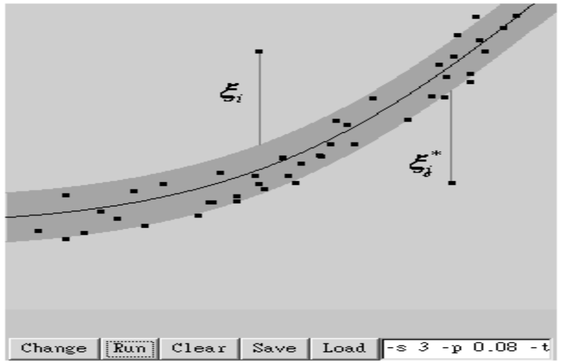
- We were given
  - Load per half hour from 1997 to 1998
  - Average daily temperature from 1995 to 1998
  - List of holidays
- Goal:

  Predict daily maximal load of January 1999
- A time series prediction problem

Slide 4

Support Vector Regression (SVR)

A nonlinear approximation of given data



Slide 5

- Data mapped to a higher dimensional space
- A linear approximation function

$$f(x) = w^T\phi(x) + b$$

- ϵ-insensitive loss function

  Given $(x_1, y_1), \ldots, (x_l, y_l)$

  Most training data in the tube

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}w^T w + C\sum_{i=1}^{l}\xi_i + C\sum_{i=1}^{l}\xi_i^*$$
$$-\epsilon - \xi_i^* \le w^T\phi(x_i) + b - y_i \le \epsilon + \xi_i,$$
$$\xi_i, \xi_i^* \ge 0, i = 1, \ldots, l.$$

Slide 6

SVR for time series prediction

- Given $(\cdots, y_{t-\Delta}, \cdots, y_{t-1}, y_t, \cdots, y_l)$ as training series
- Generate training data:

  $(y_{t-\Delta}, \cdots, y_{t-1})$ as attributes (features) of x_i

  y_t as the target value y_i
- One-step ahead prediction
- Prediction:

  Starting from the last segment

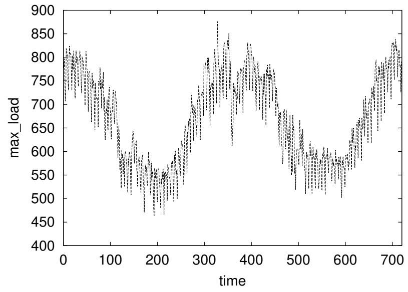
$$(y_{l-\Delta+1}, \ldots, y_l) \rightarrow \hat{y}_{l+1}$$

  Repeat by using newly predicted values

Slide 7

Data Analyses

- Maximal load of each day
  - Didn't know how to use all half-hour data
  - Not used for later analyses/experiments



Slide 8

- Issues largely discussed in earlier work
  - Seasonal periodicity
  - Weekly periodicity

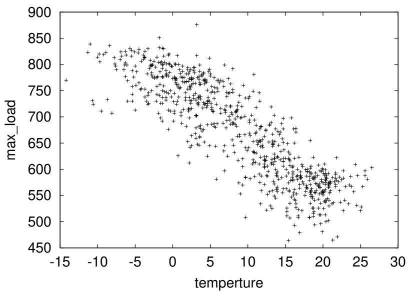
    Weekday: higher, Weekend: lower
  - Holiday effect
  - All above known for January 1999
  - Weather influence

    Temperature unknown for January 1999
- Temperature is very very important

  The main difficulty of this competition
  - Most early work on short-term prediction

    Temperature available
  - Error propagation of long-term time series prediction

Slide 9



Slide 10

Methods

- In addition to SVR, also use local modeling
- Local modeling:
  - Finding segments closely resemble the segment proceeding the point to be predicted

  - Average of elements after these similar segments of points.

Slide 11

Data Encoding

- Both method:

  Use a a segment (a vector) for predicting the next value

  Encoding: contents of a segment
- The simplest:

  Each segment: load of the previous Δ days

  Used for local model: $\Delta = 7$
- For SVM: more information incorporated
  - Seven attributes: maximal loads of the past 7 days
  - Seven binary (i.e. 0 or 1) attributes:

    target day in which day of a week

Slide 12

  - One binary attribute: target day holiday or not
  - One attribute (optional): temperature of the target day
- Temperature unknown: train two SVM

  One for load and one for temperature

$$\begin{pmatrix} y_{t-\Delta} \\ T_{t-\Delta} \end{pmatrix}, \cdots, \begin{pmatrix} y_{t-1} \\ T_{t-1} \end{pmatrix} \xrightarrow{\text{SVM1}} y_t$$

$$T_{t-\Delta}, \cdots, T_{t-1} \xrightarrow{\text{SVM2}} T_t$$

Slide 13

Implementation

- Local model: Matlab
- SVR: LIBSVM

Slide 14

LIBSVM: an Integrated Library for SVM

- Includes
  - Support vector classification/regression
  - ν-Support vector classification/regression
  - One-class support vector classification
- Features
  - Both C++/Java
  - Unbalanced data
  - Graphic interface
- Interface
  - Matlab: from Ohio State University

Slide 15

  - R interface from Technische Universität Wien
  - Python interface from HP Labs.
- Past competitions
  - Winner of OCR competition by the University of Essex and the UK post office, December 2000
  - Winner of IJCNN Challenge 2001
- Software, document, and FAQ available at
  http://www.csie.ntu.edu.tw/~cjlin/libsvm

Slide 16

Model Selection

- Parameters and features
  - Δ: for both approaches
  - Local model: # of similar segments
  - SVR:
    1. C: cost of error
    2. ϵ: width of the ϵ-insensitive tube
    3. mapping function ϕ
- Extremely important for data prediction

  Example
  - IJCNN Challenge 2001 (SVM)

Slide 17

- K. U. Leuven time series prediction competition, 1998 (local model)

- Known data separated to

  Training, validation, and testing

- January 1997 or January 1998 as testing

- Validation

  Each time one week as validation; rest as training

  Average 100 mean square errors

  Is this a robust way for time series prediction ?

- Model selection is expensive

  Restrict the search space: via reasonable choices or simply guessing

Slide 18

- $\Delta = 7$
- SVR:

  RBF function $\phi(x_i)^T\phi(x_j) = e^{-\gamma\|x_i - x_j\|^2}$

  Use default width $\epsilon = 0.5$ of LIBSVM

  C and γ: validation (average of 100 MSEs)

  $C = 2^{12}, \gamma = 2^{-4}$

Slide 19

Summer Data

- Without summer:

  Result for testing January 1998 (or 1997) better

- Give up information from April to September

- Using unsupervised time series segmentation (via SVM)

  Summer is separated

  Summer and winter should be considered separately ?

- A future issue for electricity load forecasting

Slide 20

Results: Local Model

- Validation on different number of segments

- Mainly used the closest segment

- No time to have complete tests

- Preliminary result: a little worse than SVR

Slide 21

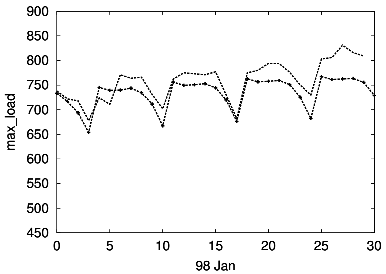
Results: SVR

- Two SVRs: very difficult to predict temperature
- One day temperature suddenly drops or increases

  ⇒ Erroneous after that day
- Conclude if temperature is used, the variation is higher
- Decide to give up using the temperature information

  Only one SVM used
- The load of each week is similar

Slide 22



- However, the model manages to find the trend
- Holiday is lower but error larger

  Results after encountering a holiday *more* inaccurate

  Holidays: January 1 and 6
- Treat all 31 days in January 1999 as non-holidays

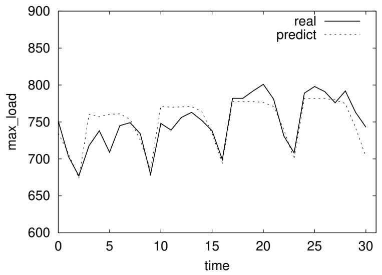
Slide 23

- Some earlier work consider holidays and non-holidays separately

  Now information about holidays is not enough.
- Overall we take a very conservative approach
- Forgot to manually lower load of January 6

  Reason why our $\max_i(\text{error}_i)$ not good
- MAPE: 1.98839%
- MSE: 364.498

Slide 24



Slide 25

Discussion

- Instead of this conservative approach, can we do better ?

  Is there a good way to use temperature information ?
- Feature selection is the key for our approach

  Example: removing summer data, treating holidays as non-holidays
- Parameter selection: needed but a large range is ok

  That is, if $C = 2^{12}, \gamma = 2^{-4}$ becomes $C = 2^{12}, \gamma = 2^{-5}$

  not much difference