60017 PERFORMANCE ENGINEERING

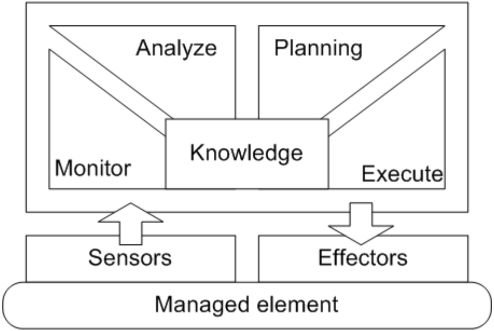
Forecasting

This lecture

- Workload forecasting
- Autoregressive models

Autoscaling system architecture

Proactive autoscaling often based on a MAPE-K control loop:



- Monitor: collect sensor data to determine symptoms
- Analyze: analyse symptoms and request a change
- Planning: plan action workflow to apply change
- Execute: enact change on the resources
- Knowledge: data shared across the control functions

Workload forecasting

MAPE implementations vary, but workload forecasting in the Analysis stage seldom changes. How does it work?

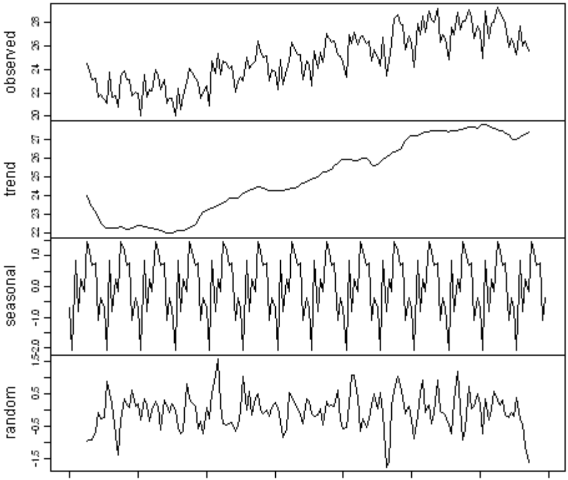
Given an arbitrary time series A_t (e.g., number of job arrived in the last period), we may decompose it as

$$A_t = T_t + S_t + I_t, \quad t = 0, 1, \ldots$$

- T_t (trend component): long-term trend (deterministic)
- S_t (seasonal component): periodic changes (deterministic)
- I_t (random component): irregular component (stochastic)

The main challenge for forecasting is to predict the random component. Detrending and deseasonalization techniques exist to expose this component from the data for model fitting.

Time series decomposition



Source: Google images.

Autoregressive models

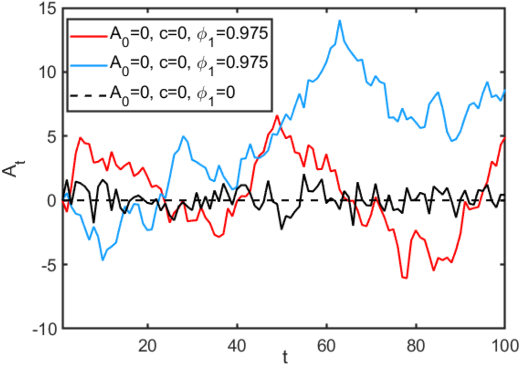
- Popular to forecast the random component of the time series
- An autoregressive model of order 1 (AR(1)) characterizes a time-series as a stochastic difference equation

$$A_t = c + \phi_1 A_{t-1} + \epsilon_t \qquad t = 1, \dots$$

  where
  - A_t is a random variable for the value at time t, A_0 is given
  - ϕ_1 and c are deterministic model parameters
  - ϵ_t is white noise, i.e., uncorrelated random variables with zero mean and finite variance ($E[\epsilon_t] = 0$, $Var[\epsilon_t] = \sigma_\epsilon^2 < +\infty$).

Example: AR(1) realizations

The $\phi_1 A_{t-1}$ term gives serial correlation to the process:



Serial correlation, in essence, gives "memory" to the process.

Forecasting

- At time t, given the knowledge of A_t, we may forecast A_{t+1} using its expected value

$$
\begin{aligned}
E[A_{t+1}|A_t] &= E[c|A_t] + \phi_1 E[A_t|A_t] + E[\epsilon_t|A_t] \\
&= c + \phi_1 A_t
\end{aligned}
$$

  where we used that $E[\epsilon_t|A_t] = E[\epsilon_t] = 0$.
- Indeed, the error between the prediction and the actual value is then

$$
A_{t+1} - E[A_{t+1}|A_t] = -\epsilon_t
$$

  which has zero mean.

Fitting the AR(1) model to the data

We can fit an AR(1) using moment matching on three statistics:

- $E[A_t] = \mu_t$: mean of the time series at time t
- $Var[A_t] = E[(A_t - \mu_t)(A_t - \mu_t)]$: variance at time t
- $K_{1,t} = E[(A_t - \mu_t)(A_{t-1} - \mu_{t-1})]$: lag-1 autocovariance at time t (a measure of serial correlation)

AR(1) fitting assumes that the time-series is stationary, meaning that the moments do not depend on t, i.e., $E[A_t] = \mu$, $Var(A_t) = V$, $K_{1,t} = K_1$, for all t.

Output: AR(1) model parameters (c, ϕ_1, σ_ϵ^2) are then obtained by solving the system of equations

$$\mu = \frac{c}{1 - \phi_1}, \quad V = \frac{\sigma_\epsilon^2}{1 - \phi_1^2}, \quad K_1 = V\phi$$

Proof: formula the AR(1) mean

We first note that

$$E[A_t] = E[c + \phi_1 A_{t-1} + \epsilon_t] = c + \phi_1 E[A_{t-1}]$$

since $E[\epsilon_t] = 0$ by definition of white noise.

Using in the stationarity assumption, we set $E[A_t] = E[A_{t-1}] = \mu$

$$\mu = c + \phi_1 \mu \Rightarrow \mu = \frac{c}{1 - \phi_1}$$

(Variance and auto-covariance formulas follow with similar passages and are left as an exercise.)

Beyond the AR(1) model

- Autoregressive model of order p (AR(p)):

$$A_t = c + \sum_{j=1}^{p} \phi_j A_{t-j} + \epsilon_t \qquad t = 1, \ldots$$

- Moving average model of order q (MA(q)):

$$A_t = \mu + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \qquad t = 1, \ldots$$

- Autoregressive moving average model (ARMA(p,q)):

$$A_t = c + \sum_{j=1}^{p} \phi_j A_{t-j} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \qquad t = 1, \ldots$$

More parameters means that the models can be more accurate, but become harder to fit.

Further extensions can also handle non-stationary time series, e.g., Autoregressive integrated moving average (ARIMA) processes.