Bayesian Merging Of Multiple Advices And Its Application To A Cold Rolling Mill

Václav Šmídl & Pavel Ettler

ÚTIA & Compureg

CSKI and DAR seminar, 15 January 2008

Content

Problem formulation

- 18 models

  system model – static / dynamic
  user target – maximum / moving average / estimated
  design method – academic / industrial / simultaneous

- Data were recorded in 6 month, i.e. many different working conditions.
- The operator **did not** follow the advices.
- The designed advisers are not directly comparable (missing variances).
- We seek an external measure of quality of advising.

Possible approaches

1. Strictly scientific experiment:
    - test all methods under exactly the same conditions.
    - impossible in production line in a factory.
2. Detailed modeling:
    - model discrepancy between the observed data and observation that would be observed if the operator followed our recommendation.
    - too much uncertainty to be modeled.
3. High-level black-box modeling
    - we build a simple auto-regressive model of the relation between two key quantities:

    a) closeness, C, of recommendations to true actions,

    b) quality of operator performance, P.

High-level model

- From a full set of 40 variables we pick two: input deviation, h_1, and output deviation, h_2.
- Large data-records are split into blocks of 1000 samples.
- Operator performance index for one block:

$$P = \frac{\mathrm{E}(h_2^2)}{\mathrm{E}((h_1 - \overline{h_1})^2)},$$

  E denotes empirical expected value on the block of data.
- Closeness of advices:

$$C_{i,t} = E\left(1 - \frac{\max\left(|u_t - u^{\star}_{i,t}|, u_t\right)}{u_t}\right),$$

  $u^{\star}_{i,t}$ is the recommended action of the ith adviser, and u_t is the actual realization.
- Lets assume that P_t is related to C_i via an unknown function, $P_t = g_i(C_i)$.

High-level model

- Lets assume that P_t is related to C_i via an unknown function,

$$P_t = g_i(C_i).$$

- Taylor expansion at operating point $\overline{C}_{i,t}$ at time t yields

$$P_t = g_i(\overline{C}_{i,t}) + g_i'(\overline{C}_{i,t})(C_{i,t} - \overline{C}_{i,t}) + e_t, \tag{1}$$

  where $g_i'()$ denotes the first derivative of $g_i()$, $\overline{C}_{i,t}$ is the fixed point of expansion, and e_t is an aggregation of higher order term.

Model: motivated by (1)

$$P_t = b_{i,t} + a_{i,t}C_{i,t} + \sigma_{i,t}v_t, \tag{2}$$

where $b_{i,t}, a_{i,t}, \sigma_{i,t}$ are unknown time-variant parameters. $v_t \sim \mathcal{N}(0,1)$ is Gaussian noise.

- time-variant parameters accommodate for time-varying expansion point, allowing fitting of the linearization to the current situation.
- Model (2) can be estimated exactly using Bayesian theory.

Merging of advices

Task:

Recommend an action, which if followed would lead to the highest operator's performance.

- Decision-making problem.
- Operator's performance is modeled by the high-level models.

Merging of advices

Task:

Recommend an action, which if followed would lead to the highest operator's performance.

- Decision-making problem.
- Operator's performance is modeled by the high-level models.

Formally:

$$u_{t+1}^{\mathrm{mer}} = \arg\min_{u_t} E(P_{t+1}|u_{t+1}).$$

$$E(P_{t+1}|u_{t+1}) = \sum_{i=1}^{18} \alpha_{i,t} f(P_{t+1}|C_{i,t+1}(u_{t+1})),$$

$$\alpha_{i,t} = f(i_t = i|P_t, C_t) \propto f(P_t|C_{i,t}, i).$$

Approximate merging

Evaluation of the formal problem is computationally prohibitive. We tested the following approximation:

1. winner takes all.

$$\alpha_{i,t} \approx [0, \ldots, 1, \ldots 0].$$
$$\hat{i} = \arg\max f(P_t | C_{i,t}, i).$$

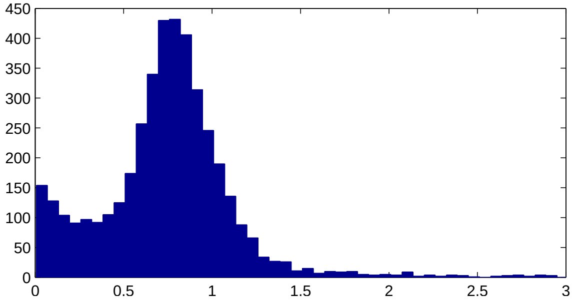
   (Choosing just one component from the mixture).

2. Avoiding optimization of $C_{i,t}(u_t)$.
   Each adviser has already designed its optimal strategy $u_{i,t}^{(o)}$, i.e.

$$u_{t+1}^{\mathrm{mer}} = \arg\min_{u_t} E(P_{t+1} | u_{t+1}).$$
$$\approx u_{\hat{i},t+1}^{(o)}.$$

Data for the experiment

- Data set collected during 6 month of production of a cold rolling mill,
- more than 4,2 million of 10 dimensional data records,
- The set contains data from a wide range of operating condition such as different materials or different passes though the mill.
- The quality of final product was within the required range for great majority of the data, and so was the operator's performance index:

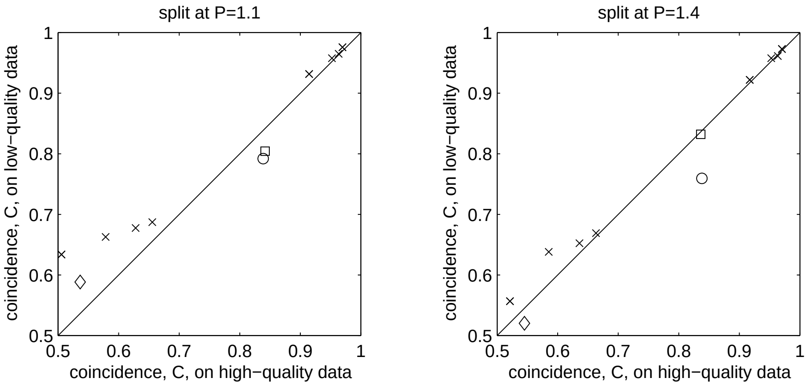


- This implies that the AGC low-level controllers worked very well, leaving only a narrow margin for improvement.

Experimental results

- Both operator's performance index and coincidence was computed for each model for each of the 4227 data batches.
- Scatter plots of these quantities form irregular clusters, discouraging visual inspection and parametric modeling of the relation.
- Hence, we propose to choose a threshold $\hat{P}$ of 'good' performance and split all data records in two sets:
  - high-quality data, $P < \hat{P}$,
  - low-quality data, $P \geq \hat{P}$.
- The rationale is that good adviser should recommend actions that are:
  - close to the actual actions when the performance is good,
  - far from the actual actions when the performance is bad.

Experimental results



- × original advisers with $C > 0.5$.
- ○ the merging adviser.
- □ adviser $M_I^{\mathrm{stat}}, M_T^{\mathrm{mov}}, M_A^{\mathrm{ind}}$.
- ◇ adviser $M_I^{\mathrm{dyn}}, M_T^{\mathrm{max}}, M_A^{\mathrm{simult}}$.

Part II Summary

- High-level black-box model was chosen as a representation of quality of advising.
- Parameters of the model were estimated using Bayesian theory.
- Merging of advices was formulated as an optimization problem under uncertainty, which was further approximated.
- The resulting algorithm is relatively robust to tuning knobs in the choice evaluation criteria.