Coordinated Development of Smart City by Intelligent Modeling and Optimization Methods

Prof. WU Cheng
CAE, Tsinghua University
2013.9.12

Contents

1. Background & Motivation

Lots of problems caused by China's rapid economic development

Serious environmental and resource problems

Social justice

Disparity between rich and poor

…

China accounted for 7 in the 10 most polluted cities of the world announced by UN

320 million people are under the unsafety of drinking because of 70% polluted rivers in China

The cost caused by pollution is about 10% of China's GDP

The consensus of the public and the government is that the economic growth must be coordinated with environment and resources

City Coordinated Development must be realized

1. Background & Motivation

Chinese government use generally kinds of evaluation index systems, which are mainly qualitative (based on statistic data), to assess a city development level

e.g.

- Developmental mode evaluation index system
- Urban and rural balanced development evaluation index system
- New urbanization comprehensive evaluation index system
- Cultural Development Index (CDI) Evaluation System

The above evaluation index systems have some shortcomings

- Afterward assessment
- The assessing subjects are mainly local cities / regions, and the impact on high-level decision is limited
- The evaluation methods are basically qualitative, and the related weights are determined by experience

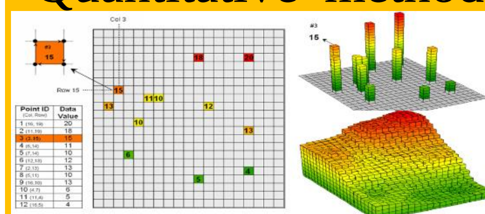
1. Background & Motivation

Can we provide a kind of methods which

Combination

Qualitative methods

Quantitative methods



to make city development planning and operational management more scientific and more coordinated

1. Background & Motivation

Adopting qualitative/quantitative combination-based intelligent modeling methods for predicting the key city development & operation indicators which is concerned by mayors (e.g., GDP, fiscal revenue, energy consumption, environmental protection)

Adopting model, qualitative/quantitative-based intelligent optimization decision methods for optimizing the related decisions that can impact the above indicators

Provide an intelligent optimization decision tool for the mayors and leaders of key municipal sectors

Smart city optimization decisions

- Help to solve the most-concerned city decision problems
- Make the city development & operation more scientifically and rationally

Contents

中国工程院
Chinese Academy of Engineering

2.1 Various problems in city development & operation

Various problems are involved in city development & operation

- Economic growth(GDP growth)
- Fiscal revenue growth
- Per capital income growth
- Employment growth
- Public service improvement
- Resource (such as energy) constraint
- Emission constraint
- City carrying capacity
- …

These problems need to be considered and solved

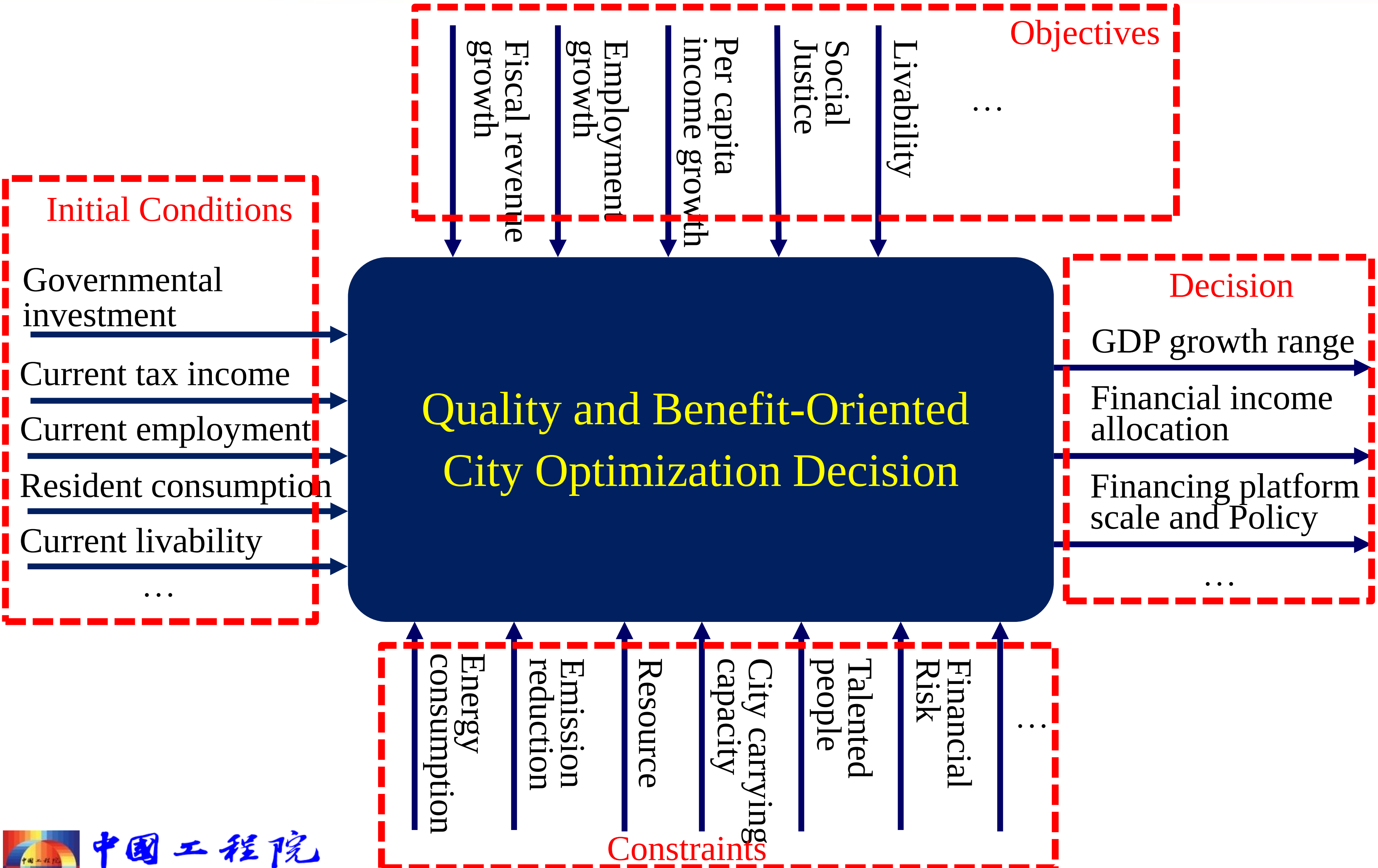
✓ coordinately

✓ to satisfy the public opinion and emergencies

2.2 Complexity and solving of the problem

Difficulties	Solutions
City development is a gaming result between market (invisible hand) and government regulation (visible hand)	■ Adopt intelligent models for predicting the market trends ■ Take government regulations as decision variables
Scientific description is very difficult because the above problem may be semi-structured or unstructured	Adopt intelligent modeling and optimization algorithms which can handle semi-structured and unstructured data
The city operational data may not real and not accurate	■ Take the city which operational data are relative accurate as a pilot ■ Adopt the uncertainty method (e.g., fuzzy theory) to handle the uncertainty data
Public opinion can influence government decision-making	Adopt the unstructured method to describe the public opinion
Emergencies	Adopt dynamic optimization mechanism

2.3 We choose 'Quality and Benefit–Oriented City Optimization Decision'



2.4 Problem Solving Process

The optimization problem for the city development & operation can be concluded as followings

Based on initial conditions

- Past government investment
- Past fiscal revenue
- Past employment
- Past per capita incomes
- Past resident consumption
- Current livability level

Satisfying kinds of constraints

- Energy Consumption
- Emission
- Resource (water etc..)
- City carrying capacity
- Talented people
- Financial risk

Make related decisions

- GDP growth
- Government revenue allocation
- Financing platform scale

Realize this year's targets

- Fiscal revenue growth
- Employment growth
- Per capita income growth
- Social justice indicators
- Livability indicators

A series models should be created based on historical data

2.5 Formal Description of Smart City Coordinated Development

max F, max M, max C, min E, max P, max L, max S

- F - Fiscal revenue growth
- M - Employment growth
- C - Per capita income growth
- E - Energy consumption
- P - Pollution emission
- L - Livability
- S - Social Justice

… ….

s.t.

Range constraints of decision variables $D_l \leq D(such\ as\ GDP\ growth, \ldots) \leq D_u$

Equality constraints of decision variables $\varphi(D) = D_e$

Inequality constraints of decision variables $\emptyset(D) \leq D_i$

where D are the decision variables which determine the objectives F, M, C, E and etc.

2.6 Formal Description of Smart City Coordinated Development (cont.)

Prediction model plays an important role in city optimization decision

Two relationships are very important

Indicator Model

Decision (regulation mean) → City development & operation indicator

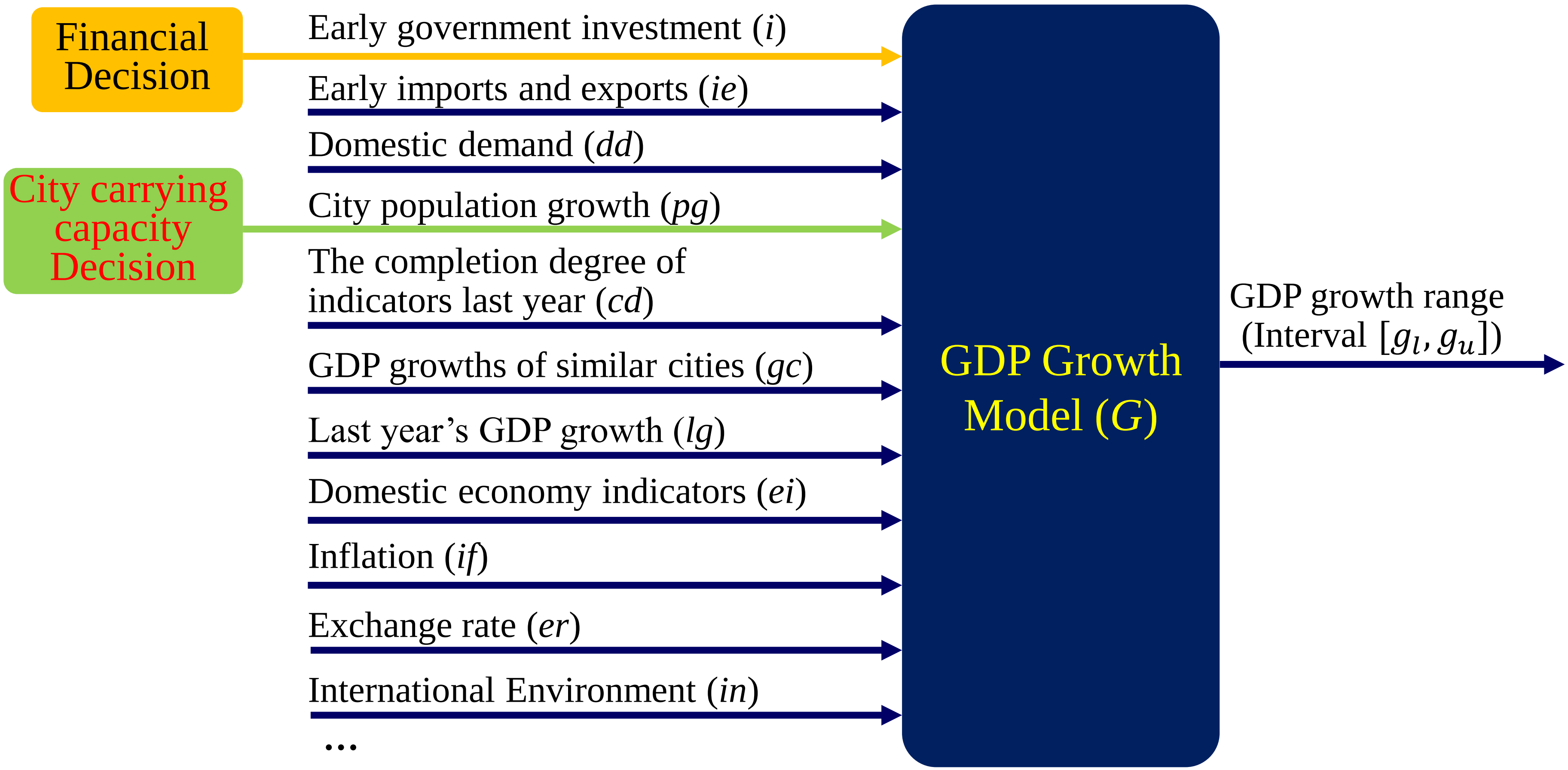
e.g., the relationship between GDP growth and fiscal revenue, employment rate, per capital income, energy consumption and etc.

Decision Constraint Model

Related city operational information → Range constraint of decision

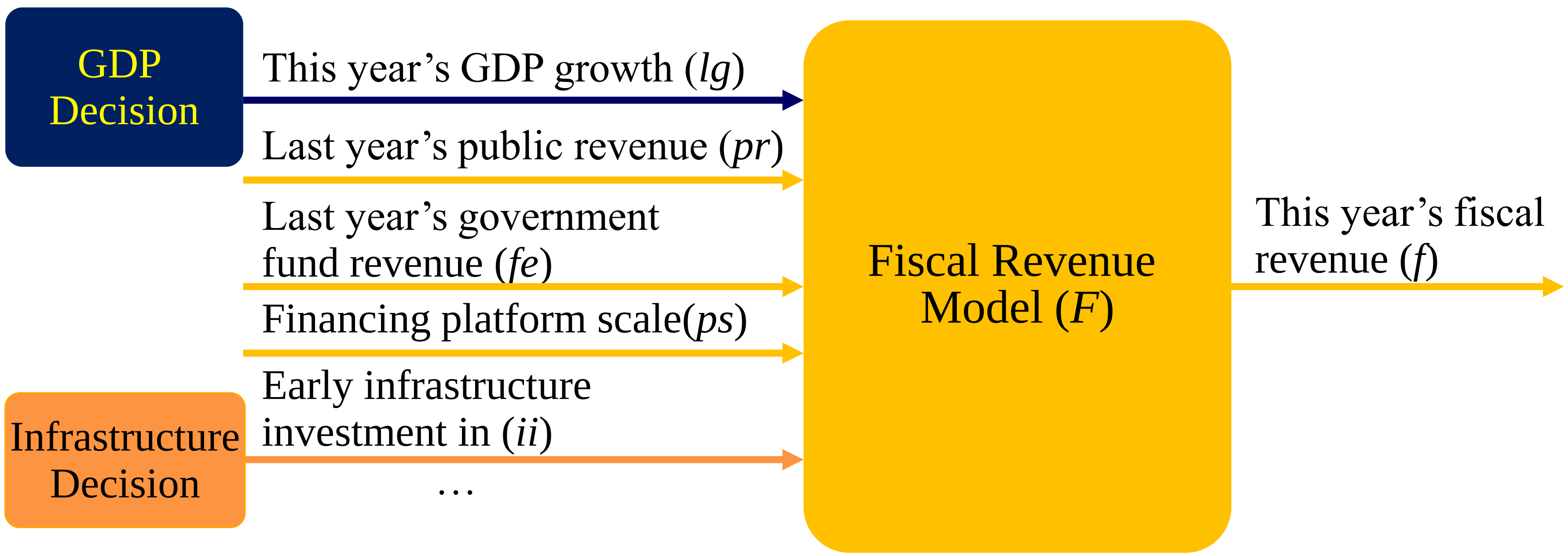
e.g., the variable range of GDP growth is constrained by past investment, the developments of surrounding cities, national policies and etc.

2.7 Prediction model Examples: GDP Growth Model



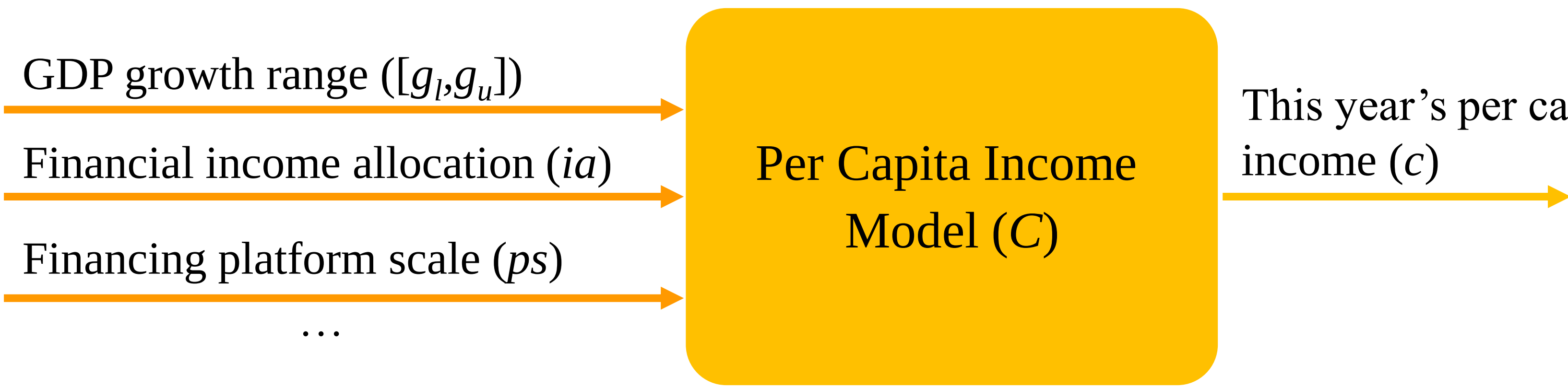
$$[g_l, g_u] = G(i, ie, dd, pg, cd, gc, lg, ei, if, er, in, \ldots)$$

2.7 Prediction model Examples: Fiscal Revenue Model



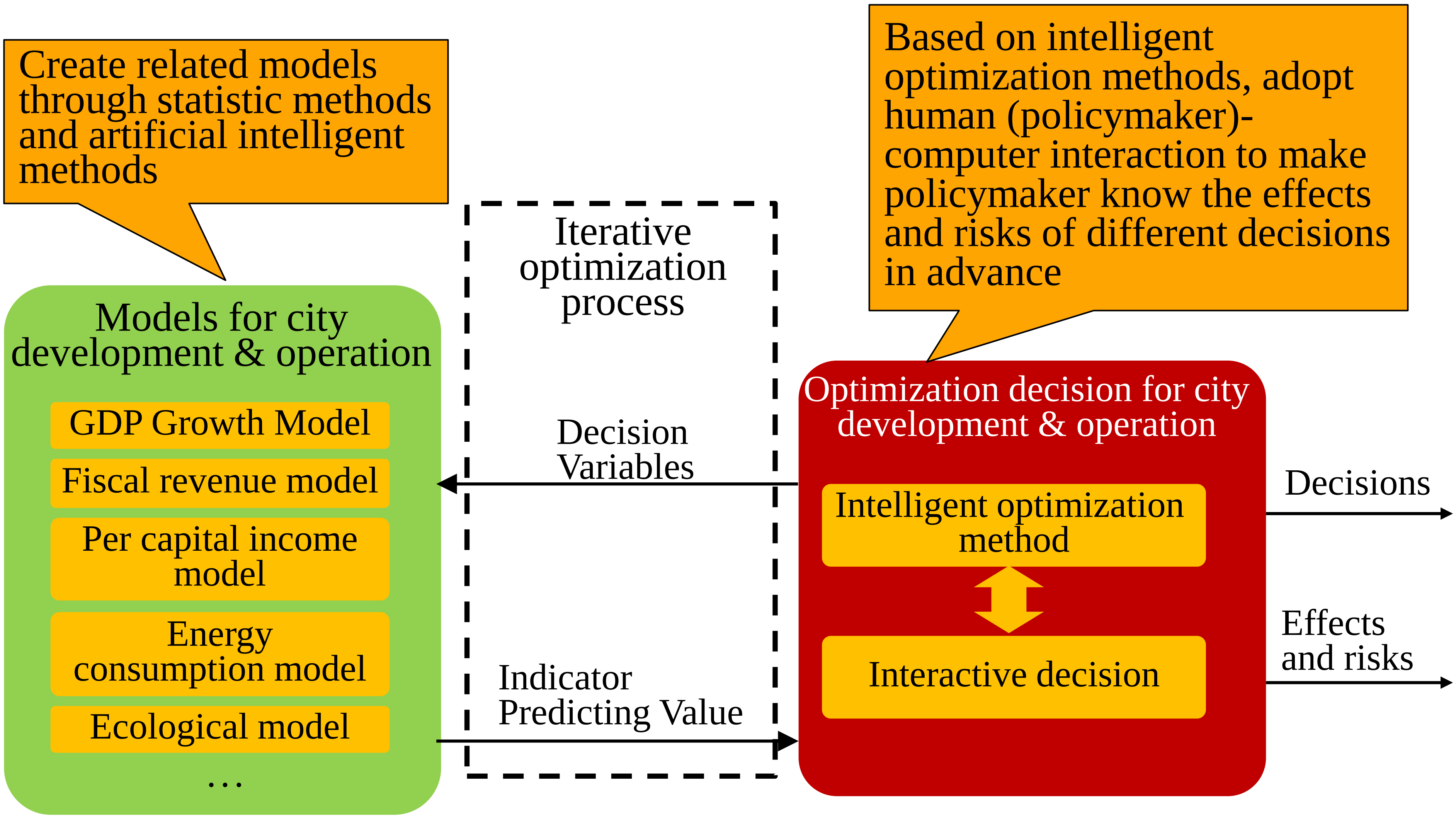
$$f = F(lg, pr, fe, ps, ii, \dots)$$

2.7 Prediction model Examples: Per Capita Income Model



$$c = C([g_l, g_u], ia, ps, \dots)$$

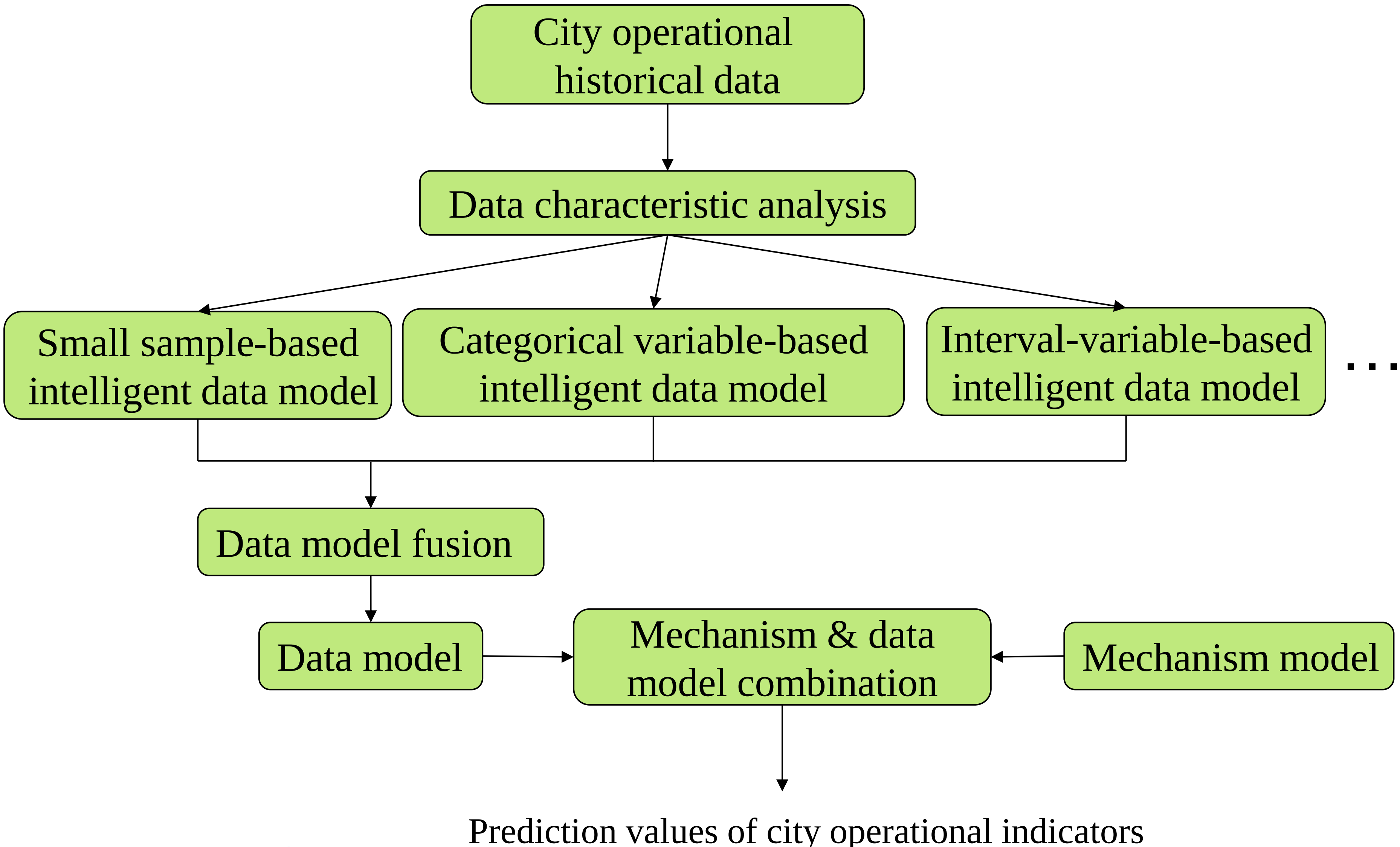
2.8 Overall Solution of Intelligent Modeling and Optimization for City Coordinated Development



Contents

中国工程院
Chinese Academy of Engineering

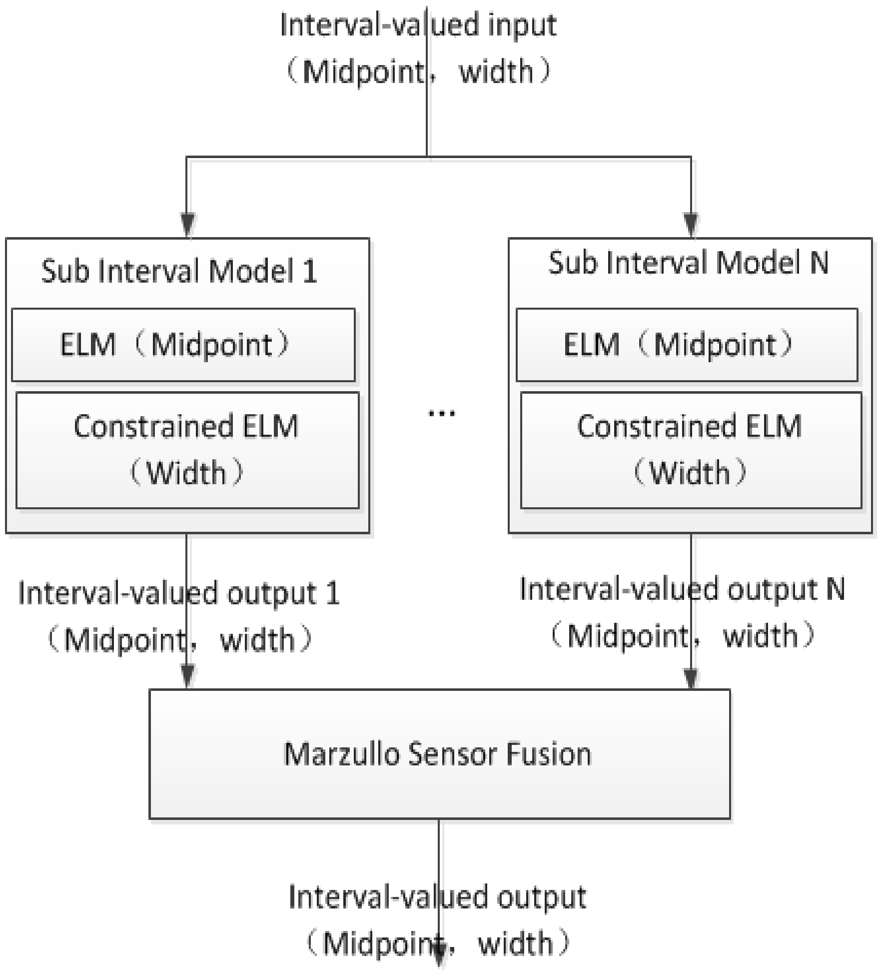
3.1 Mechanism & data combination-based Intelligent Modeling Framework



a) Midpoint and width-based Interval ELM modeling method

Some city indicator models which input/output variables are interval, such as GDP growth

- A new interval-value-based ELM (I-ELM) is used to construct interval nonlinear models for crisp or interval-valued input data and interval-valued output data
- Corresponding to the two characteristics of interval value (i.e., midpoint and width), I-ELM includes two parts:
  - ✓ An ordinary ELM for modeling the midpoint
  - ✓ An constraint ELM for modeling the width
- To improve the prediction accuracy, Marzullo sensor fusion algorithm is adopted



3.2 Interval-variable-based intelligent data model
a) Midpoint and width-based Interval ELM modeling method

Midpoint Model

- The ordinary ELM is used for modeling the midpoint

$$\mathbf{T}^M = \mathbf{H}^M \boldsymbol{\beta}^M + \boldsymbol{\varepsilon}^M$$

- The estimated value of parameters to be determined in the above model

$$\hat{\boldsymbol{\beta}}^M = ((\mathbf{H}^M)^T \mathbf{H}^M)^{-1} (\mathbf{H}^M)^T \mathbf{T}^M$$

Width Model

- Interval-valued data T must have $\mathbf{T}^U \geq \mathbf{T}^L$, which introduces the model should ensure the predicted value $\hat{\mathbf{T}}^W = \mathbf{H}^W \hat{\boldsymbol{\beta}}^W \geq \mathbf{0}$. So, the width model becomes **a constrained ELM**

$$\mathbf{T}^W = \mathbf{H}^W \boldsymbol{\beta}^W + \boldsymbol{\varepsilon}^W$$

$$s.t.\ \tilde{\mathbf{H}}^W \boldsymbol{\beta}^W \geq \mathbf{0}$$

Add constraints to the parameters to be learned.

- The above problem can be solved using the **constrained least-squares estimation**

$$\hat{\boldsymbol{\beta}}^W = ((\mathbf{H}^W)^T \mathbf{H}^W)^{-1} (\mathbf{H}^W)^T \mathbf{T}^W + ((\mathbf{H}^W)^T \mathbf{H}^W)^{-1} (\tilde{\mathbf{H}}^W)^T \boldsymbol{\lambda}^*$$

Interval-value fusion

- Generate the midpoint & the width model n_f times. Suppose there is an input datum$\{(\mathbf{x}_{N+1}^{M}, \mathbf{x}_{N+1}^{W}) | \mathbf{x}_{N+1}^{M} \in \mathbf{R}^{n_1}, \mathbf{x}_{N+1}^{W} \in \mathbf{R}^{n_2}\}$, the predicted interval-values $(\hat{t}_{N+1,k}^{L}, \hat{t}_{N+1,k}^{U}), k = 1, \ldots, n_f$, can be calculated as follows:

$$\hat{t}_{N+1,k}^{L} = \hat{t}_{N+1,k}^{M} - \hat{t}_{N+1,k}^{W} / 2 = H_{N+1,k}^{M} \hat{\beta}_{k}^{M} - H_{N+1,k}^{W} \hat{\beta}_{k}^{W} / 2$$

$$\hat{t}_{N+1,k}^{U} = \hat{t}_{N+1,k}^{M} + \hat{t}_{N+1,k}^{W} / 2 = H_{N+1,k}^{M} \hat{\beta}_{k}^{M} + H_{N+1,k}^{W} \hat{\beta}_{k}^{W} / 2 。$$

The output of the sub-interval models.

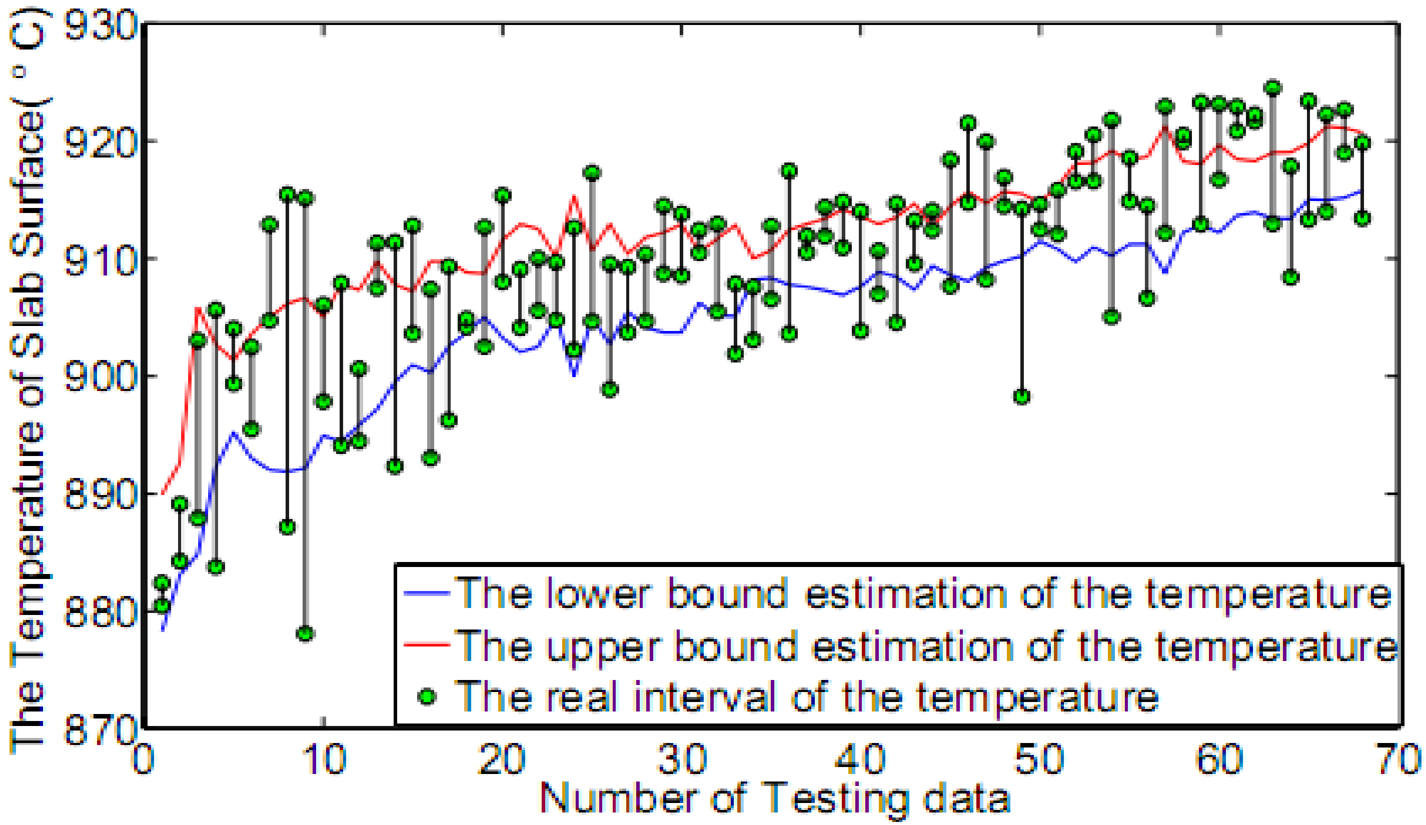
- Suppose t is the number of the faulty output values, and n_f -t is the number of the correct output values. Using the Marzullo Sensor Fusion algorithm, we can get

$$\hat{t}_{i_1, i_2, \ldots, i_{(n_f - j)}} = \bigcap_{k=1}^{n_f - j} \hat{t}_{i_k}, \; j = 0, \ldots, t, i_1, i_2, \ldots, i_{(n_f - j)} \in \{1, 2, \ldots, n_f\} \qquad S_{n_f - j} = \bigcup_{\hat{t}_{i_1, i_2, \ldots, i_{n_f - j}} \neq \varnothing} \hat{t}_{i_1, i_2, \ldots, i_{n_f - j}}$$

- Find the minimum value $\hat{t}_{N+1,fusion}^{L}$ in set S and the maximum value $\hat{t}_{N+1,fusion}^{U}$. Then, the final output result should be the interval $\hat{t}_{N+1,fusion} = [\hat{t}_{N+1,fusion}^{L}, \hat{t}_{N+1,fusion}^{U}]$.

3.2 Interval-variable-based intelligent data model
a) Midpoint and width-based Interval ELM modeling method

Predicting the temperature of slab surface in continuous casting process



b) Asymmetric Gaussian Bayesian and ELM-based modeling method

Overview

- ✓ This method is more accurate and more fast modeling efficiency than the above method
- ✓ Adopt asymmetric Gaussian Bayesian and ELM (Extreme Learning Machine)-mixed learning method
- ✓ Obtain upper bound and lower bound models through the adaptive modification of a pair of reciprocal weights

Modeling problem description

- ✓ Suppose the input dimension of training sample data is N, and its output dimension is 1, then we have

$$\{(\mathbf{x}_i, t_i)\}_{i=1}^{N};$$

$$\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,n}), i = 1, \ldots, N$$

- ✓ Introduce the asymmetric Gaussian distribution into ELM, then

$$p(t \mid \boldsymbol{\beta}, b, w) = \sqrt{\frac{2b}{\pi}} \frac{1}{w+1} \begin{cases} e^{-\frac{b}{2}(t-\mathbf{h}\boldsymbol{\beta})^2} & \text{if } \ t < \mathbf{h}\boldsymbol{\beta}, \\ e^{-\frac{b}{2\mathrm{w}^2}(t-\mathbf{h}\boldsymbol{\beta})^2} & \text{otherwise.} \end{cases}$$

Modeling problem description (cont.)

- ✓ Then, the likelihood function of sample data is

$$p(\mathbf{t}\,|\,\boldsymbol{\beta},b,w)=\left(\sqrt{\frac{2b}{\pi}}\frac{1}{w+1}\right)^{N}e^{-\frac{b}{2}\left(\|\mathbf{t}_1-\mathbf{H}_1\boldsymbol{\beta}\|^2)+\|\mathbf{t}_2-\mathbf{H}_2\boldsymbol{\beta}\|^2/w^2)\right)}$$

- ✓ Use Gaussian prior distribution for β (the output layer weight of ELM), that is

$$p(\boldsymbol{\beta}\,|\,a)=\left(\frac{\mathrm{a}}{2\pi}\right)^{\frac{M}{2}}\prod_{k=1}^{M}e^{-\frac{a}{2}\beta_k^2}$$

- ✓ Then, we can obtain the Asymmetric Gaussian Bayesian ELM

$$p(\boldsymbol{\beta}_1\,|\,\mathbf{t})=\frac{p(\mathbf{t}\,|\,\boldsymbol{\beta}_1,b_1,w_1)\,p(\boldsymbol{\beta}_1\,|\,a_1)}{p(\mathbf{t})}\qquad p(\boldsymbol{\beta}_2\,|\,\mathbf{t})=\frac{p(\mathbf{t}\,|\,\boldsymbol{\beta}_2,b_2,w_2)\,p(\boldsymbol{\beta}_2\,|\,a_2)}{p(\mathbf{t})}$$

$$p(\mathbf{t}\,|\,a_1,b_1)=\int p(\mathbf{t}\,|\,\boldsymbol{\beta}_1,b_1,w_1)\,p(\boldsymbol{\beta}_1\,|\,a_1)d\boldsymbol{\beta}_1\qquad p(\mathbf{t}\,|\,a_2,b_2)=\int p(\mathbf{t}\,|\,\boldsymbol{\beta}_2,b_2,w_2)\,p(\boldsymbol{\beta}_2\,|\,a_2)d\boldsymbol{\beta}_2$$

where w_1, w_2 are a pair of reciprocal

The learning process of the new ELM

- Use Bayesian theory, $p(\beta|t)$ can be described as followings

$$p(\boldsymbol{\beta}\,|\,\mathbf{t}) = \frac{p(\mathbf{t}\,|\,\boldsymbol{\beta},b,w)\,p(\boldsymbol{\beta}\,|\,a)}{p(\mathbf{t})} = \frac{(w+1)^{-N}\,2^{\frac{N-M}{2}}\,\pi^{-\frac{N+M}{2}}\,b^{\frac{N}{2}}\,a^{\frac{M}{2}}\,e^{-\mathbf{M}(\boldsymbol{\beta})}}{p(\mathbf{t})}$$

- Let $\dfrac{\partial \ln p(\boldsymbol{\beta}\,|\,\mathbf{t})}{\partial \boldsymbol{\beta}} = 0$, we have $\hat{\boldsymbol{\beta}} = b\mathbf{C}^{-1}(\mathbf{H}_1^T\mathbf{t}_1 + \dfrac{1}{w^2}\mathbf{H}_2^T\mathbf{t}_2)$

- Then, we have

$$\begin{aligned}
p(\mathbf{t}\,|\,a,b) &= \int p(\mathbf{t}\,|\,\boldsymbol{\beta},b,w)\,p(\boldsymbol{\beta}\,|\,a)\,d\boldsymbol{\beta} \\
&= (w+1)^{-N}\,2^{\frac{N-M}{2}}\,\pi^{-\frac{N+M}{2}}\,b^{\frac{N}{2}}\,a^{\frac{M}{2}} \times \int e^{-\mathbf{M}(\boldsymbol{\beta})}\,d\boldsymbol{\beta} \\
&= (w+1)^{-N}\,2^{\frac{N-M}{2}}\,\pi^{-\frac{N+M}{2}}\,b^{\frac{N}{2}}\,a^{\frac{M}{2}} \times e^{-\mathbf{M}(\hat{\boldsymbol{\beta}})} \int e^{-\frac{1}{2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T\mathbf{C}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})}\,d\boldsymbol{\beta} \\
&= (w+1)^{-N}\,2^{\frac{N-M}{2}}\,\pi^{-\frac{N+M}{2}}\,b^{\frac{N}{2}}\,a^{\frac{M}{2}} \times e^{-\mathbf{M}(\hat{\boldsymbol{\beta}})}\,(2\pi)^{\frac{N}{2}}\,|\mathbf{C}|^{-\frac{1}{2}}
\end{aligned}$$

■ The learning process of the new ELM (cont.)

- ✓ Let

$$\frac{\partial \ln p(\mathbf{t} \mid a, b)}{\partial a} = \frac{M}{2a} + \frac{\|\boldsymbol{\beta}\|^2}{2} - \frac{\mathrm{tr}(\mathbf{C}^{-1})}{2} = 0$$

- ✓ Finally, we obtain

$$a = \frac{M}{\|\boldsymbol{\beta}\|^2 + \mathrm{tr}(\mathbf{C}^{-1})}$$

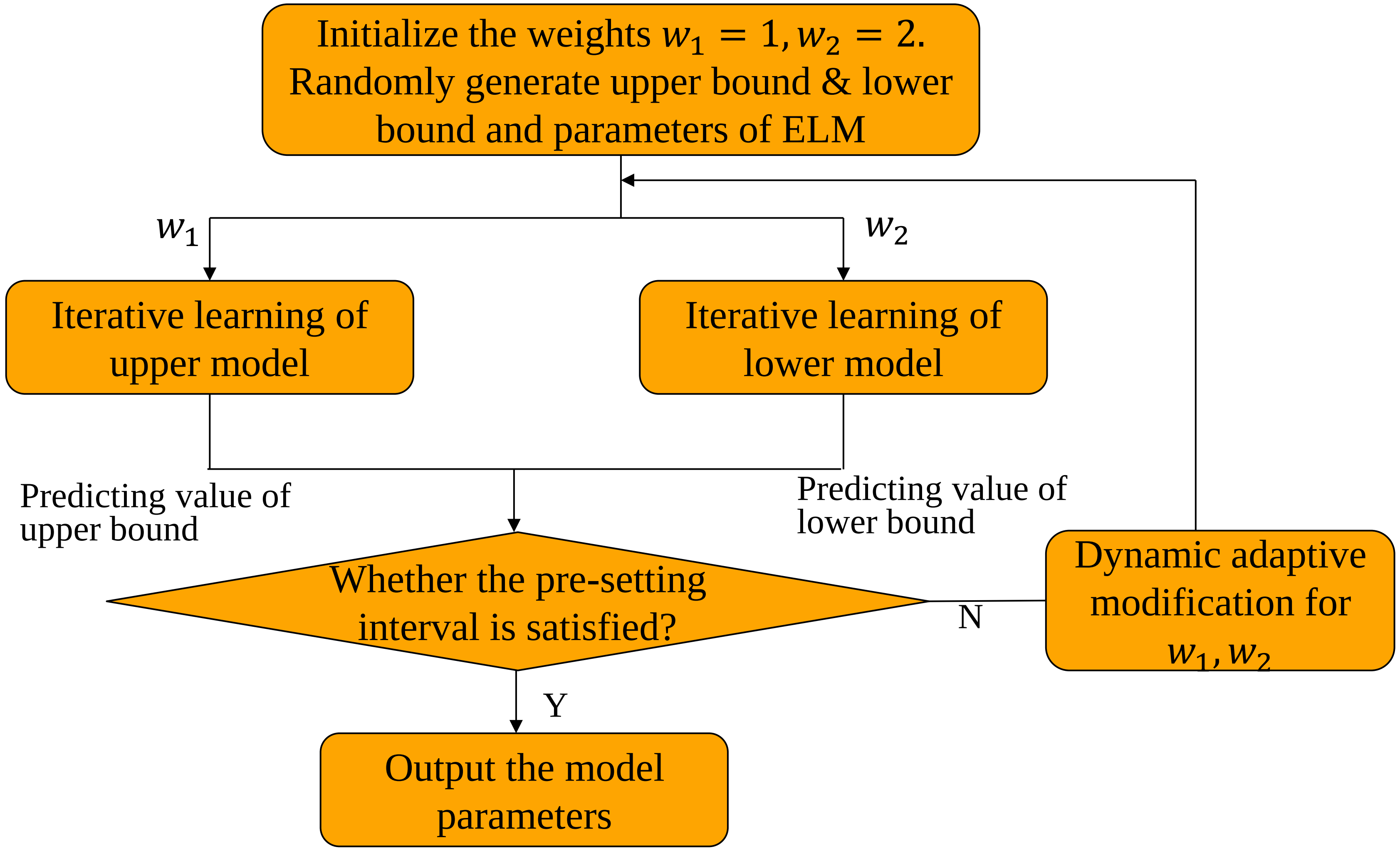
- ✓ Similar, let $\frac{\partial \ln p(\mathbf{t} \mid a, b)}{\partial b} = \frac{N-\gamma}{2b} - \frac{\|\mathbf{t}_1 - \mathbf{H}_1\boldsymbol{\beta}\|^2}{2} - \frac{\|\mathbf{t}_2 - \mathbf{H}_2\boldsymbol{\beta}\|^2}{2w^2} = 0$

- ✓ We obtain

$$\mathrm{b} = \frac{N-\gamma}{\|\mathbf{t}_1 - \mathbf{H}_1\boldsymbol{\beta}\|^2 + \|\mathbf{t}_2 - \mathbf{H}_2\boldsymbol{\beta}\|^2 / w^2}$$

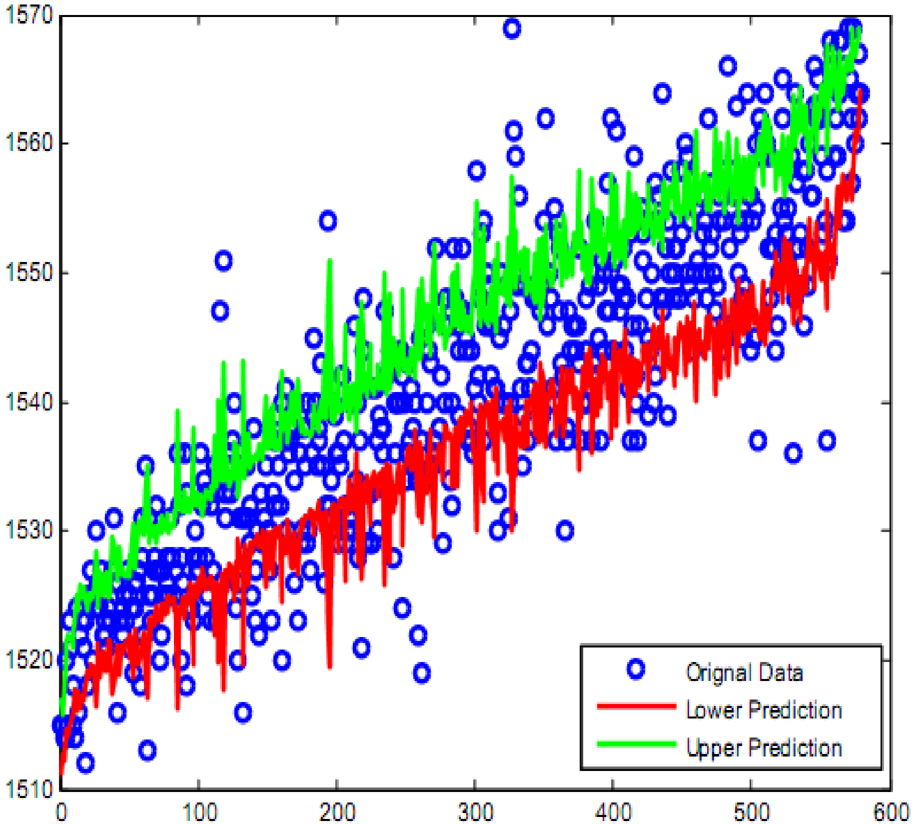
b) Asymmetric gaussian Bayesian and ELM-based modeling method

■ Algorithm process

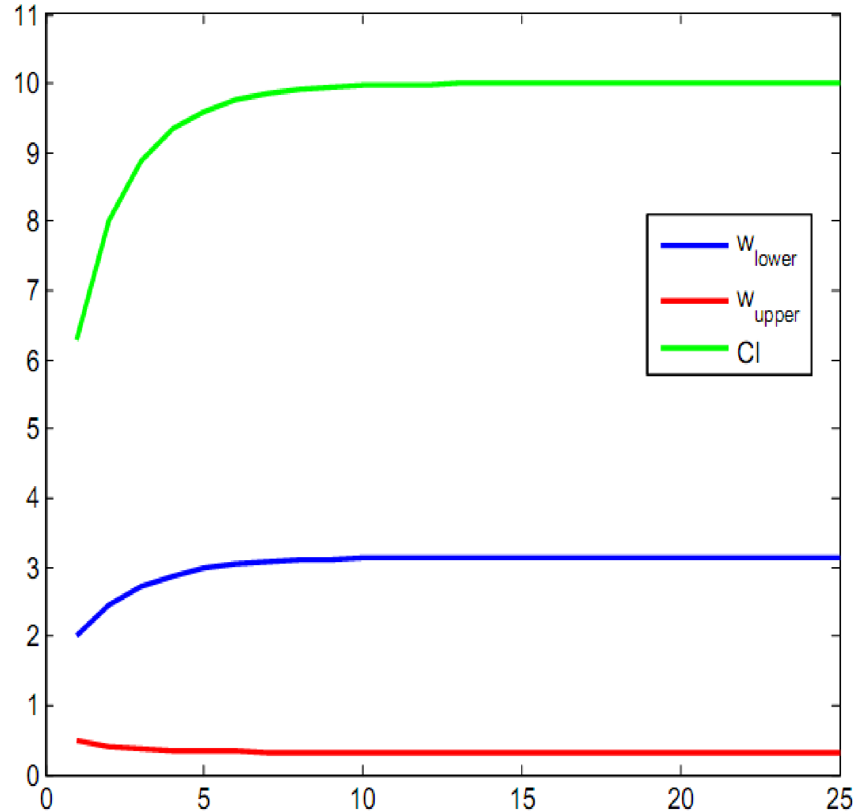


■ Numerical experiments

Predicting the temperature of molten steel



Predicting performance

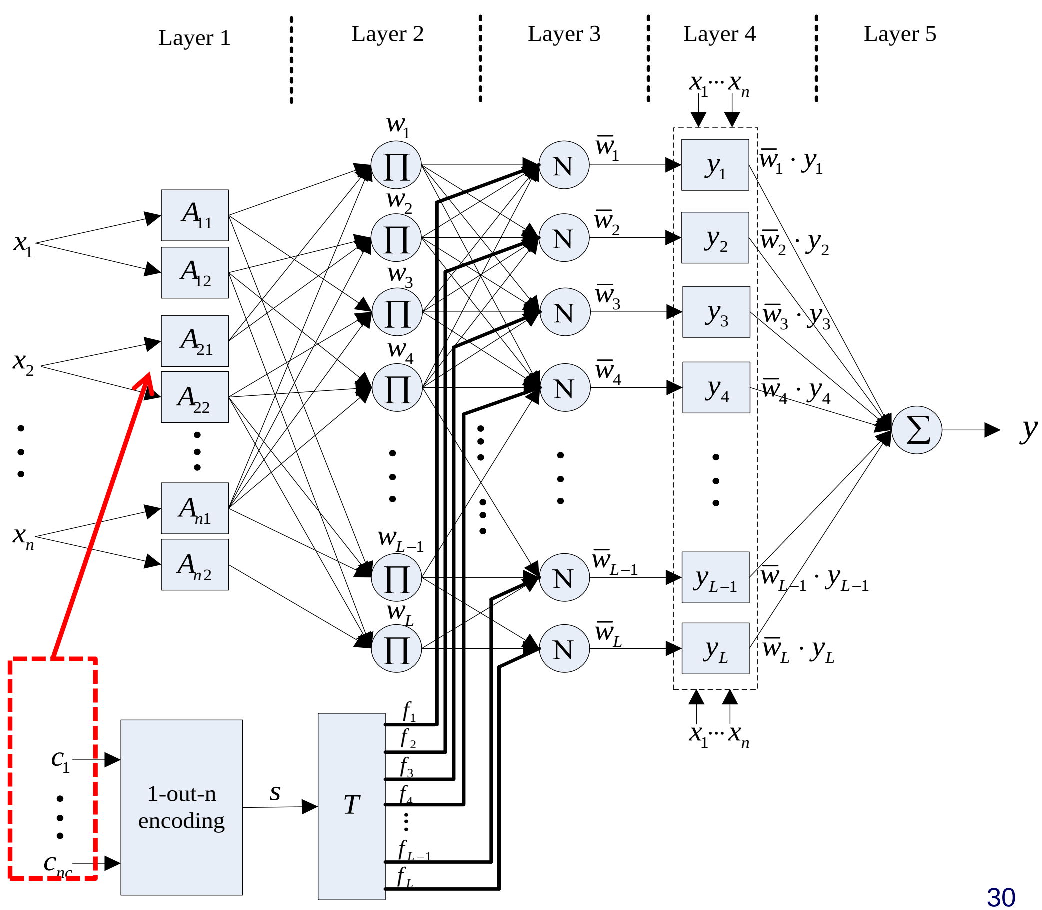


The adjusting curve of the weights

3.4 A new ANFIS (Adaptive Neuro-network Fuzzy Inference System) modeling method with categorical inputs

Aiming at the categorical variables in some model's inputs, such as the scale of financial platform could be big, relatively big, normal, relatively small and small

- A firing-strength-matrix is introduced in ANFIS
- Its output is connected directly to the Layer 3 of ANFIS
- The new ANFIS structure can handle categorical variables

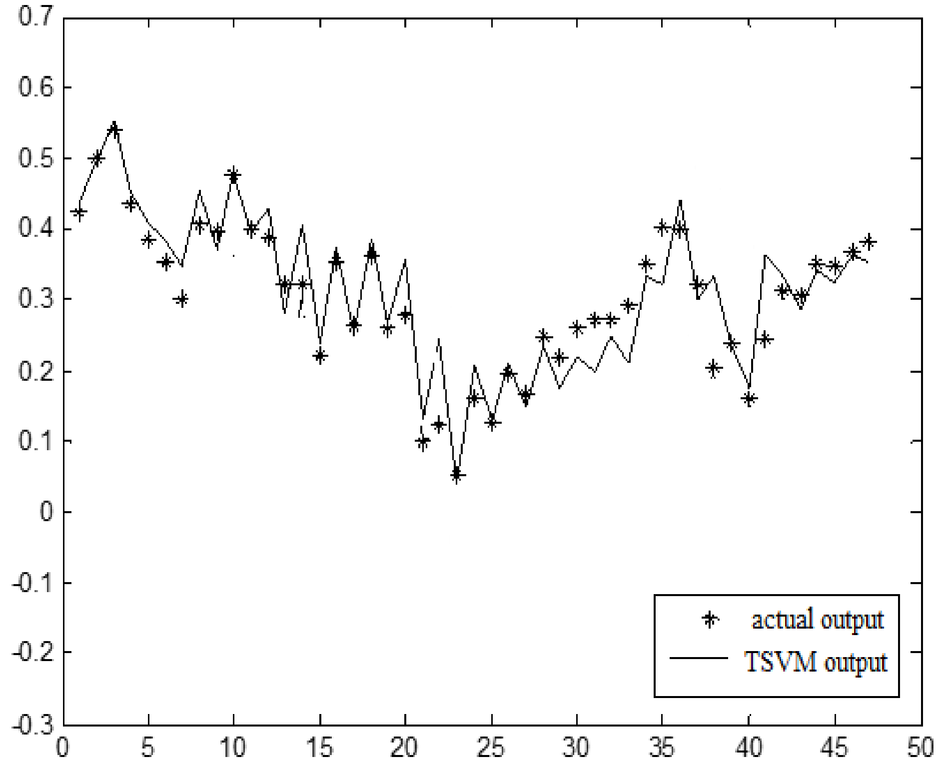


3.5 residual entropy-based two layer SVM (Support Vector Machine)

Aiming at the lack of data, such as some city operational data are obtained monthly or quarterly

- Use residual entropy to measure the residual information in training errors, and create two layer SVM
- When a new data is arrived, adopt the KKT condition of the quadratic optimization in outer SVM to determine how to modify the outer SVM with this data
- After the training errors of outer SVM are obtained, Use the residual entropy to measure the degree of certainty of training errors of outer SVM
- When the residual entropy is big, the inner SVM is created using the above training errors
  - ✓ If the inner SVM was already created, use the new data and prediction error of outer SVM to adjust the inner SVM

Temperature prediction of the molten glass in direction-changing interval of glass furnace



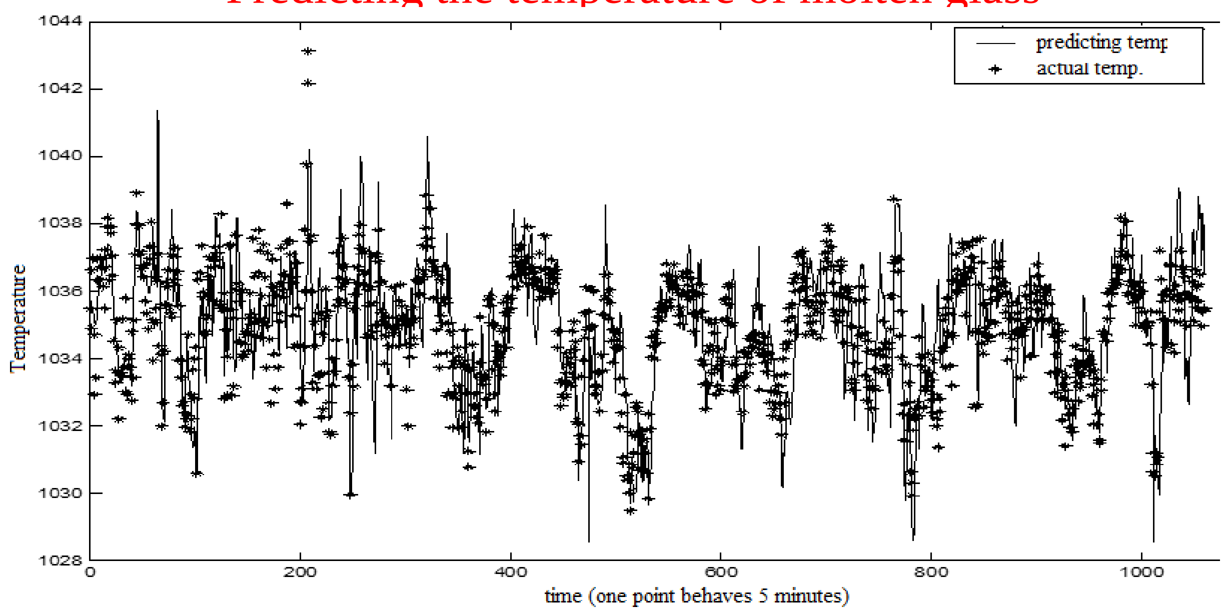
The experimental results show that this method outperforms normal SVM method by 5.7422% on average

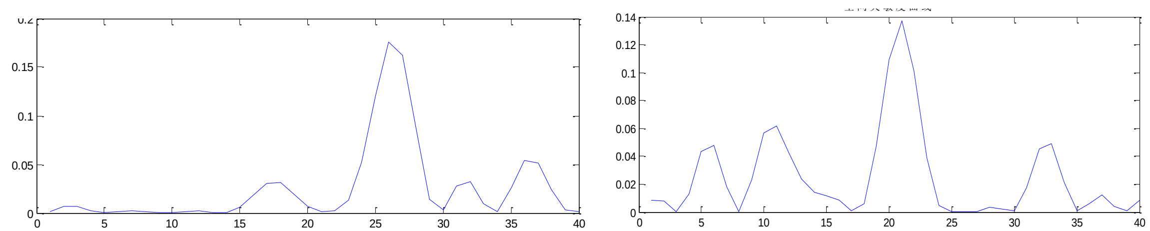
3.6 A process-neural-network and sensitivity analysis-based modeling method

Aiming at the multiple uncertain delays in some indicator model and city operational data, such as when the policy and investment could influence the city operation are not uncertain

- Sensitivity analysis for the weights between input layer and hidden layer on the time dimension
- Multiple-uncertain-lag matrix is introduced, and through this matrix the unrelated time intervals of input are filtered out
- This method don't determine the time lags explicitly

Predicting the temperature of molten glass





These are the curves of sensitivity analysis. The larger the sensitivity, the more significant influence on output of this time interval.

Contents

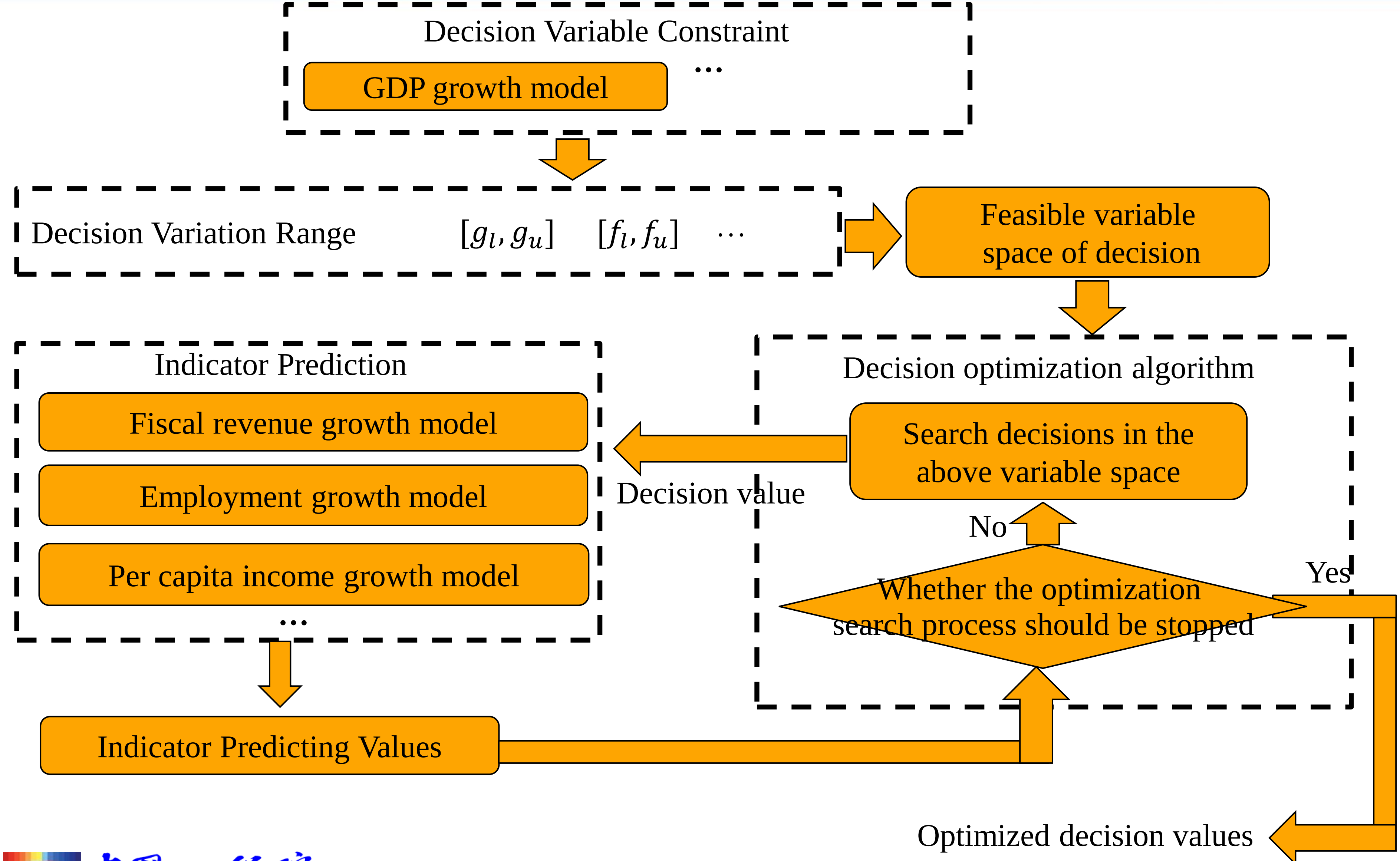
4, Intelligent Optimization Method

After creating the above models, adopt intelligent optimization method to determine the decisions for city development & operation

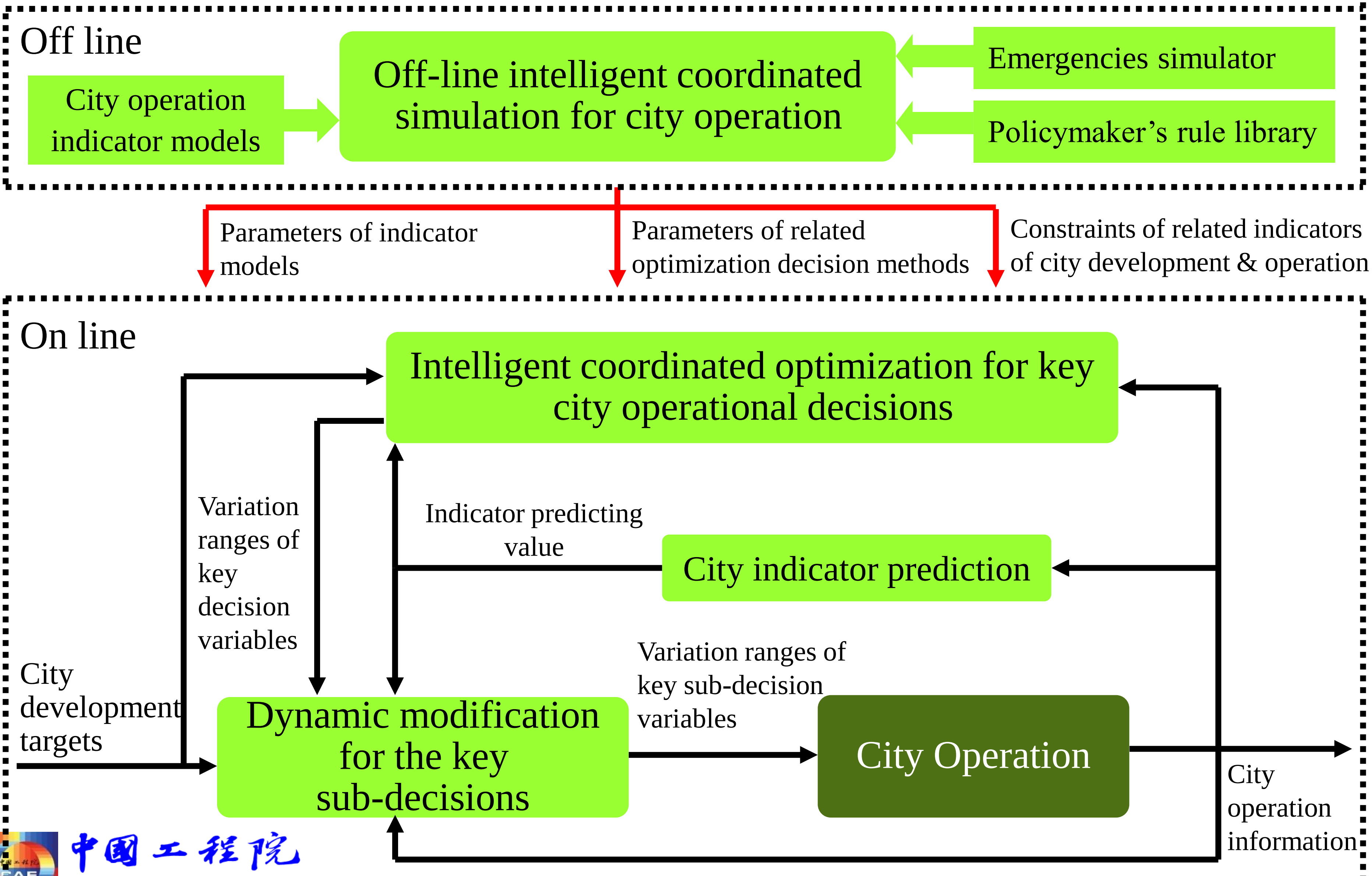
Difficulties

Large scale	e.g., the fiscal income allocation decision involves kinds of decision variables (investments), such as infrastructure, medicine, investment, education, environment protection and etc.
Strong constraints	e.g., the energy consumption and pollution emission are both rigid targets
Multi-objectives	e.g., the coordination between city development and environmental pollution, the coordination between per capita income and city infrastructure investment
Uncertainty	e.g., many uncertainty factors and emergencies influence the city operation

Overview of City Decision Optimization Process

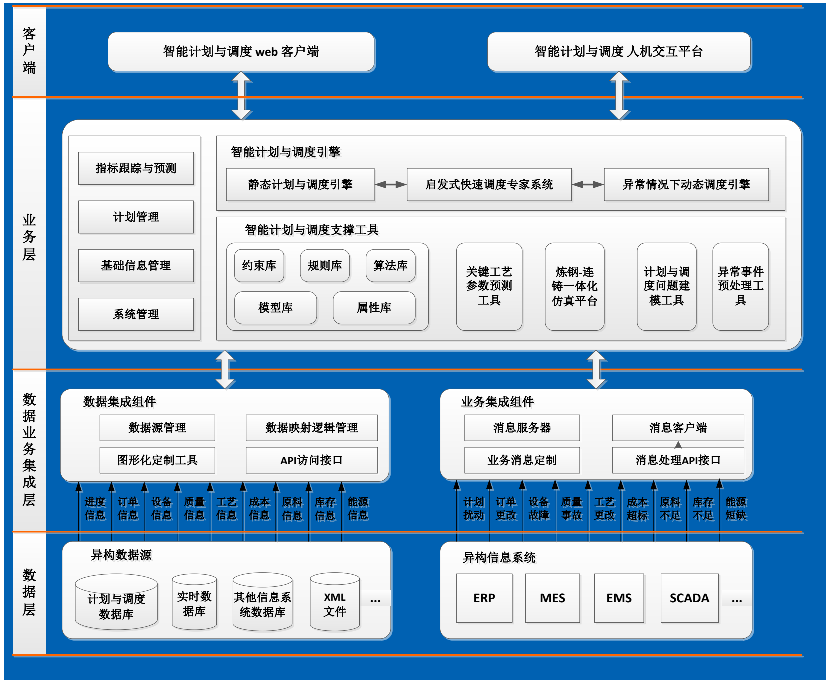


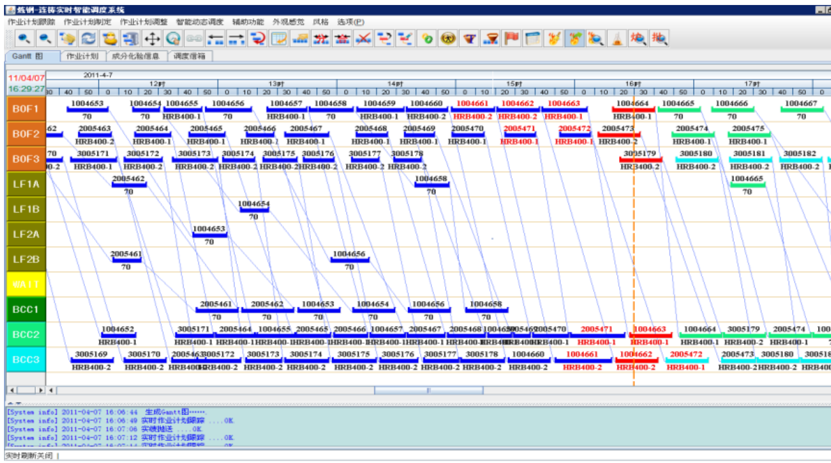
Intelligent Optimization Decision Framework for city development and operation



As Demo

- Shagang Group from Jiangsu Province is
  - ✓ the 2nd from Mainland China in the “most competitive world-class iron and steel enterprises” published by WSD*
  - ✓ the 366th in the Top 500 Enterprises of the world
- Applied System: Intelligent Steelmaking Planning & Scheduling System
  - ✓ Intelligent & automatic steelmaking planning
  - ✓ Production index forecasting
  - ✓ Plan execution & feedback
  - ✓ Intelligent dynamic scheduling
- **Significant application effect (2011)**
  - ✓ **Output is increased by 4%, added output value is 0.4B RMB, added profit is 37M RMB**
  - ✓ **Average machine utilization in continuous-casting is improved by 6.81%**
  - ✓ **Power consumption in refine furnace per ton is decreased by 11.69%**
  - ✓ **Energy consumption per ton is decreased by 36%**





CSMC

A LEADING ANALOG FOUNDRY

- CSMC pioneered the open foundry business model in China since 1997 in China, which capacity has attained 110,000 wafers per month
- Proposed solution: Intelligent Scheduling & Operation Optimization Integration Solution for Large-scale Semi-conductor Manufacturing (with corresponding system)
- Applied region: diffusion region of CSMC

- Based on carrying forward the advanced concept of lean production, achievements via our system is as followings

index	before implementation	after implementation
Completion Ratio of Month Plan	100%	105.50%
Cycle Time of NONE DMOS	1.91 day/layer	1.69 day/layer
Cycle Time of DMOS	4.26 day/layer	3.73 day/layer

Contents

- Background & Motivation
- Basic Models and Formal Descriptions of Smart City Coordinated Development
- Intelligent Modeling Methods
- Intelligent Optimization Methods
- Conclusion

5, Conclusion

- City development should be coordinated, especially for smart cities.
- There are many challenges related to coordinated problem. Intelligent methods provide a promising way.
- The proposed intelligent modeling and optimization methods have been applied to several enterprises. They are corrective、effective and valued for complex problem solving.
- Quantitative analysis of city coordinated development is an open research question. It needs more research efforts and applications.

Thank you !