

Identifikacija dinamičnega sistema z znanim modelom šuma z modelom na osnovi Gaussovih procesov

Kristjan Ažman[†], Juš Kocijan^{†,‡}

[†]Inštitut Jožef Stefan, Ljubljana

[‡]Univerza v Novi Gorici, Nova Gorica

kristjan.azman@ijs.si

Identification of the dynamic system with known noise model using Gaussian process model

Dynamic system identification with Gaussian process prior model is an emerging, probabilistic, non-parametric modelling method for identification. An advantage of this model is the possibility to include various prior knowledge into the model. An example of such prior knowledge is a known noise model of the system's output. In this paper the incorporation of the prior knowledge about the noise model in the form of ARMA model will be presented together with an illustrative dynamic system identification example.

1 Uvod

Za identifikacijo linearnih dinamičnih sistemov [5] obstaja veliko metod, identifikacija nelinearnih dinamičnih sistemov pa je v splošnem kompleksnejša. V prispevku bo predstavljena identifikacija dinamičnih sistemov z modelom na osnovi Gaussovih procesov (GP modelom), ki spada med verjetnostne neparametrične modele. Izhod iz GP modela je normalna (Gaussova) porazdelitev, opisana s srednjo vrednostjo in varianco. Srednja vrednost predstavlja najbolj verjetno vrednost izhoda, varianca pa mero zaupanja v to napoved. Poleg variance napovedanega izhoda je dobra lastnost GP modela tudi možnost vključevanja različnih vrst predznanja kot npr. vključevanje lokalnih modelov [1], statične karakteristike, histereze, predznanja o šumu [2]. Prav vključevanje predznanja o motilnem šumu bo tema tega prispevka. Njegov doprinos je vključevanje predznanja o modelu šuma za dinamične sisteme. Do sedaj je bilo vključevanje tovrstnega predznanja prikazano le za statične sisteme.

GP model je bil najprej uporabljen za reševanje problemov regresije, popularnost v krogu ljudi, ki se ukvarjajo s strojnimi učenjem, pa je pridobil pozneje z deli Rasmussen-a [8] in Neal-a [7].

Evropski projekt MAC je spodbudil raziskovanje uporabe GP modela za identifikacijo dinamičnih sistemov, kar je botrovalo k objavi prvih prispevkov s tega področja, npr. [4]. Ker je identifikacija z GP modelom sveže področje raziskovanja, je potrebno še precej dela, da se spozna in ovrednoti njegove prednosti, med katere spada tudi vključevanje predznanja.

V prispevku bo predstavljena uporaba GP modela za identifikacijo dinamičnega sistema s znanim modelom šumnega procesa na izhodu. Delovanje GP modela bo predstavljeno na preprostem primeru dinamičnega sistema.

V nadaljevanju bo najprej predstavljen GP model, njegova uporaba za identifikacijo dinamičnih sistemov in način vključevanja predznanja o šumu v GP model. Temu bo sledil primer identifikacije nelinearnega dinamičnega sistema z GP modelom z in brez vključenega predznanja. V zaključku bo na kratko povzeto bistvo prispevka.

2 Modeliranje z modelom na osnovi Gaussovih procesov

Identifikacija z metodo na osnovi Gaussovih procesov je statistična, neparametrična metoda identifikacije. Bolj podrobno je predstavljena npr. v [9, 10].

Gaussov proces je normalno porazdeljen proces, povsem določen z vektorjem srednjih vrednosti in kovariančno matriko. Nanj lahko gledamo kot na množico naključnih spremenljivk, porazdeljenih po normalni porazdelitvi: $f(\mathbf{x}_1), \dots, f(\mathbf{x}_n) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. Elementi Σ_{ij} kovariančne matrike $\mathbf{\Sigma}$ so kovariance med vrednostmi spremenljivk $f(\mathbf{x}_i)$ in $f(\mathbf{x}_j)$ ter so funkcije argumentov \mathbf{x}_i in \mathbf{x}_j : $\Sigma_{ij} = C(\mathbf{x}_i, \mathbf{x}_j)$. Za računanje kovariance med spremenljivkami lahko uporabimo katerokoli kovariančno funkcijo $C(\mathbf{x}_i, \mathbf{x}_j)$, ki generira nenegativno kovariančno matriko $\mathbf{\Sigma}$. Pri predpostavki o stacionarnosti in gladkosti procesa navadno izberemo Gaussovo kovariančno funkcijo [9]:

$$C(\mathbf{x}_i, \mathbf{x}_j) = v \exp \left[-\frac{1}{2} \sum_{d=1}^D w_d (x_i^d - x_j^d)^2 \right] \quad (1)$$

kjer je D velikost vhodnega vektorja \mathbf{x} in je $\Theta = [w_1, \dots, w_D, v]^T$ vektor parametrov, imenovanih hiperparametri. Tako se imenujejo, ker določajo verjetnostno porazdelitev funkcij. Hiperparameter v določa velikost kovariance, hiperparametri w_i pa izražajo relativno pomembnost posameznih komponent x^d vhodnega vektorja \mathbf{x} .

Vzemimo sistem:

$$y(k) = f(\mathbf{x}(k)) + \epsilon(k), \quad (2)$$

kjer je $\mathbf{x}(k)$ vhodni vektor in je $\epsilon(k)$ beli šum z varianco v_0 , $\epsilon \sim \mathcal{N}(0, v_0)$. Predpostavimo, da je so izhodi družine funkcij $f(\cdot)$ porazdeljeni normalno s kovariančno funkcijo (1). Iz tega verjetnostnega okvira sledi [9]: $y_1, \dots, y_N \sim \mathcal{N}(0, \mathbf{K})$ kjer $\mathbf{K} = \Sigma + v_0 \mathbf{I}$ z enotno matriko \mathbf{I} velikosti $N \times N$.

Na podlagi poznavanja N parov učnih podatkov $\{\mathbf{x}_i, y_i\}_{i=1}^N$ bi radi napovedali porazdelitev za y^* pri nekem novem vhodu \mathbf{x}^* . Za naključne spremenljivke (y_1, \dots, y_N, y^*) velja:

$$\begin{bmatrix} \mathbf{y} \\ y^* \end{bmatrix} \sim \mathcal{N}(0, \mathbf{K}_{N+1}) \quad (3)$$

s kovariančno matriko

$$\mathbf{K}_{N+1} = \begin{bmatrix} \mathbf{K} & \mathbf{k}(\mathbf{x}^*) \\ \mathbf{k}(\mathbf{x}^*)^T & k(\mathbf{x}^*) \end{bmatrix} \quad (4)$$

kjer je $\mathbf{y} = [y_1, \dots, y_N]^T$ vektor učnih izhodov.

To skupno verjetnostno porazdelitev lahko razdelimo na dva dela: *a-priori* verjetnost učnih podatkov $\mathbf{y}|\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$, kjer je \mathbf{X} matrika učnih vhodov velikosti $N \times D$ in pogojno verjetnostno porazdelitev izhoda y^* glede na porazdelitev učnih podatkov.

Neznane parametre kovariančne funkcije (1) in varianco šuma v_0 lahko dobimo z metodo največje podobnosti (ang. *maximum likelihood*).

Pri znanih hiperparametrih je pogojna porazdelitev (3) pri vhodu \mathbf{x}^* :

$$p(y^*|\mathbf{y}, \mathbf{X}, \mathbf{x}^*) = \frac{p(\mathbf{y}, y^*)}{p(\mathbf{y}|\mathbf{X})} \quad (5)$$

Ta porazdelitev je Gaussova [9] s srednjo vrednostjo in varianco:

$$\mu(\mathbf{x}^*) = \mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1} \mathbf{y} \quad (6)$$

$$\sigma^2(\mathbf{x}^*) = k(\mathbf{x}^*) - \mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}^*) + v_0 \quad (7)$$

kjer je $\mathbf{k}(\mathbf{x}^*) = [C(\mathbf{x}_1, \mathbf{x}^*), \dots, C(\mathbf{x}_N, \mathbf{x}^*)]^T$ vektor kovarianc med učnimi vhodi in testnim vhodom in $k(\mathbf{x}^*) = C(\mathbf{x}^*, \mathbf{x}^*)$ avtokovarianca testnega vhoda.

Področja vhodnega prostora, kjer je odvisnost podatkov kompleksna ali kjer je na voljo malo podatkov ali so močno pošumljeni, so tako opisana s povečano varianco.

2.1 Identifikacija dinamičnega sistema z GP modelom

Pri identifikaciji dinamičnega sistema kot vhod v GP model $\mathbf{x}(k)$ namesto vektorja neodvisnih spremenljivk nastopa vektor regresorjev:

$$\mathbf{x}(k) = [y(k-1), \dots, y(k-L), u(k-1), \dots, u(k-L)]^T \quad (8)$$

ki bo imel za sistem L -tega reda $2L$ komponent. Pri simulaciji takega modela kot vhode uporabljamo pretekle vrednosti napovedi modela $\hat{y}(i)$, saj pravih vrednosti izhoda $y(i)$ ne poznamo. Več o simulaciji GP modelov bralec najde v [3, 4, 1].

2.2 Vključevanje predznanja o šumu v GP model

Predhodno je bil predstavljen GP model, kjer smo na izhodu iz sistema (2) predpostavili beli Gaussov šum. V kolikor je ta šum barvni in poznamo njegov model, lahko to predznanje vključimo v kovariančno matriko [2].

Kovariančno funkcijo $C(\mathbf{x}_i, \mathbf{x}_j)$ lahko predstavimo kot:

$$C(\mathbf{x}_i, \mathbf{x}_j) = C_f(\mathbf{x}_i, \mathbf{x}_j) + C_n(\mathbf{x}_i, \mathbf{x}_j) \quad (9)$$

kjer $C_f(\cdot)$ predstavlja funkcijski del kovariance, opisan npr. z (1), $C_n(\cdot)$ pa šumni del kovariance med vhodnima vektorjema \mathbf{x}_i in \mathbf{x}_j . Pri belem Gaussovem šumu na izhodu neznanega procesa je kovarianca $C_n(\mathbf{x}_i, \mathbf{x}_j) = v_0 \delta_{ij}$ in različna od nič samo v primeru, da gre za isti vhodni vektor $\mathbf{x}_i = \mathbf{x}_j$. Pri barven šumu na izhodu pa so vrednosti šuma korelirane s preteklimi vrednostmi in kovarianca je različna od nič tudi za različne vhodne vektorje. Predstavljajmo si, da je šum na izhodu iz procesa opisan z avtoregresijskim (AR) modelom z drsečim povprečjem (MA) r -tega reda:

$$\begin{aligned} \epsilon(t) &= e(t) - a_1 \epsilon(t-1) \dots - a_r \epsilon(t-r) + \\ &+ b_1 e(t-1) + \dots + b_r e(t-r) = \frac{B(q)}{A(q)} e(t), \end{aligned} \quad (10)$$

kjer je $e(t)$ beli Gaussov šum. Označimo šumni del $C_n(\epsilon(t), \epsilon(t+\tau))$ kot $C(\tau)$.¹ Če vstavimo vhodna vektorja ϵ_t in $\epsilon_{t+\tau}$:

$$\begin{aligned} \epsilon_t &= e_t + b_1 e_{t-1} + \dots + b_r e_{t-r} - a_1 \epsilon_{t-1} - \dots - a_r \epsilon_{t-r} \\ \epsilon_{t+\tau} &= e_{t+\tau} + \dots + b_r e_{t+\tau-r} - a_1 \epsilon_{t+\tau-1} - \dots - a_r \epsilon_{t+\tau-r}, \end{aligned}$$

kjer smo časovno odvisnost zaradi poenostavitve namesto s (t) označili z indeksom t , lahko kovarianco $C(0)$ zapišemo:

$$\begin{aligned} C(0) &= C(\epsilon_t, \epsilon_t) = \sigma^2 (1 + b_1^2 + \dots + b_r^2) + \\ &C(\epsilon_t, -a_1 \epsilon_{t-1} - \dots - a_r \epsilon_{t-r}). \end{aligned} \quad (11)$$

¹V nadaljevanju bomo indeks n , ki označuje šumni del kovariančne matrike izpuščali zaradi poenostavitve zapisa.

Prvi del enačbe (11) predstavlja prispevek MA dela in ga zapišemo krajše kot B_0 , drugi del pa predstavlja prispevek AR dela. $C(0)$ lahko zapišemo kot:

$$C(0) = B_0 - a_1 C(1) - a_2 C(2) - \dots - a_r C(r), \quad (12)$$

saj $C(\tau) = C(-\tau)$. Podobno kot zgoraj lahko za $\tau \neq 0$ šumni del kovariance med dvema različnimi vhodnima vektorjema ϵ_t in $\epsilon_{t+\tau}$ zapišemo kot:

$$\begin{aligned} C(\tau) &= B_\tau - a_1 C(\tau - 1) - \dots - a_r C(\tau - r), \text{ kjer} \\ B_\tau &= \sigma^2(1 \cdot b_i + b_1 \cdot b_{i+1} + \dots + b_{r-i} \cdot b_r), \tau \leq r \\ B_\tau &= 0, \tau > r \end{aligned} \quad (13)$$

Iz prvih $r+1$ gornjih enačb dobimo *Yule-Walkerjev* sistem enačb:

$$\begin{bmatrix} 1 & a_1 & \dots & a_{r-1} & a_r \\ a_1 & 1+a_2 & \dots & a_r & 0 \\ a_2 & a_1+a_3 & \dots & 0 & 0 \\ \vdots & & & \ddots & \vdots \\ a_r & a_{r-1} & \dots & a_1 & 1 \end{bmatrix} \begin{bmatrix} C(0) \\ C(1) \\ C(2) \\ \vdots \\ C(r) \end{bmatrix} = \begin{bmatrix} B_0 \\ B_1 \\ B_2 \\ \vdots \\ B_r \end{bmatrix} \quad (14)$$

Z njegovo rešitvijo dobimo prvih $r+1$ elementov kovariančne matrike $C(0) \dots C(r)$. Preostale elemente kovariančne matrike $C(\tau)$, $\tau = r+1 \dots N-1$ dobimo iterativno z rešitvijo enačbe:

$$C(i) = -a_1 C(i-1) - a_2 C(i-2) - \dots - a_r C(i-r) \quad (15)$$

V kovariančno matriko \mathbf{K} vstavimo kovariance med posameznimi vhodnimi vektorji \mathbf{x}_i in \mathbf{x}_j kot:

$$C(\mathbf{x}_i, \mathbf{x}_j) = C_f(\mathbf{x}_i, \mathbf{x}_j) + C_n(|i-j|) \quad (16)$$

3 Primer vključevanja predznanja o šumu v GP model

3.1 Primer dinamičnega sistema

Kot primer za prikaz prednosti, ki jih prinaša vključevanje predznanja o šumu v GP model, smo si izbrali nelinearni dinamični sistem [6]:

$$y(k+1) = \frac{y(k)}{1+y^2(k)} + u^3(k) + \epsilon(k) \quad (17)$$

kjer barvni šum $\epsilon(k)$ z varianco $v_0 = 0.01$ opišemo z ARMA modelom s parametri $a_1 = -0.6$, $a_2 = 0.08$, $b_0 = 1$ in $b_1 = b_2 = 0$.

Naš cilj je bila izgradnja GP modela, ki bo opisoval dinamično obnašanje sistema (17) na področju, omejenem z vhomom u med $u_{\min} = 0$ in $u_{\max} = 1.5$. Za ovrednotenje dobljenega modela smo primerjali rezultate simulacije GP modela z obnašanjem sistema. Uporabili smo dva različna GP modela:

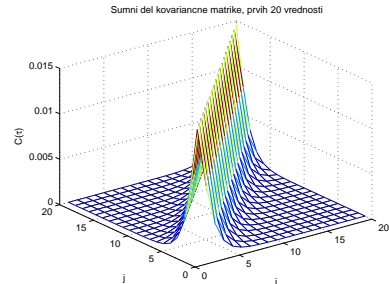
- GP model brez predznanja o šumu in
- GP model z vključenim predznanjem o modelu šuma na izhodu.

3.2 Učenje GP modelov

Za učenje GP modela smo uporabili vzorce, dobljene iz dinamičnega odziva modela. Ker nas je zanimalo, kako se odraža vključevanje predznanja o šumu, smo uporabili zaporedne vzorce vhodnega in izhodnega signala, kjer je korelacija šuma med vzorci največja. Za vzbujanje sistema je bil uporabljen psevdo-naključni binarni signal (PNBS) [5] s taktom $k_{sw} = 2$, kjer amplituda ob spremembi lahko zavzame katerokoli vrednost med u_{\min} in u_{\max} .

Najprej je bil narejen GP model za sistem (17) brez vključevanja znanja o šumnem procesu v GP model. Obnašanje sistema je opisovalo 202 vzorcev, ki smo jih uporabili za učenje. V k -tem koraku sta vhodni vektor v GP model sestavljala vhod in izhod iz sistema $[u(k-1), y(k-1)]$, učni izhod GP modela pa izhod $y(k)$. Ker predznanja o šumu nismo uporabili, smo na izhodu predpostavili beli šum in optimizirali hiperparametre v, w_1, w_2 ter v_0 .

Za učenje drugega modela, t.j. GP modela z vključenim znanjem o šumu na izhodu procesa, so bili uporabljeni isti vzorci kot za učenje prvega modela. Predznanje o šumu smo opisali s šumnim delom kovariančne matrike \mathbf{K}_n kot je bilo opisano v prejšnjem razdelku. Vidimo ga na sliki 1. Ker smo šumni proces poznali, nam hiperparametra v_0 ni bilo potrebno optimirati.



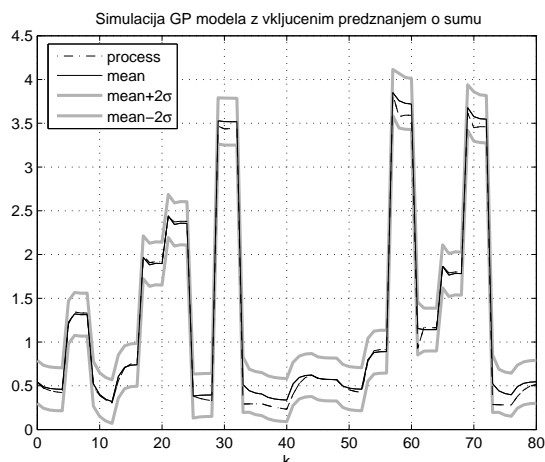
Slika 1: Šumni del kovariančne matrike \mathbf{K}_n

3.3 Rezultati

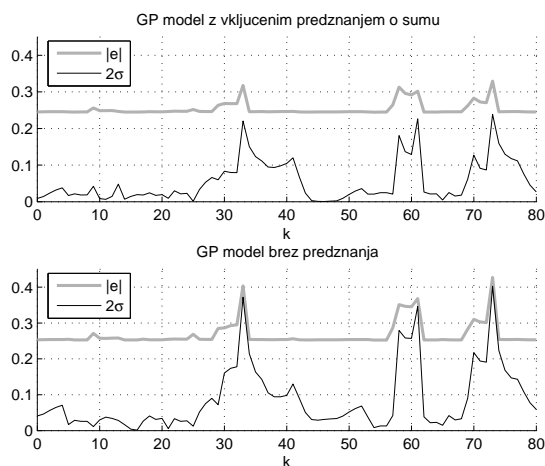
Za ovrednotenje obeh modelov smo modela simulirali s PNBS signalom s taktom 4 in primerjali rezultate simulacije z odzivom sistema. Rezultate ovrednotenja modela z vključenim predznanjem vidimo na sliki 2. Na sliki 3 je primerjava med napako GP modelov z in brez vključenega predznanja o modelu šuma. Zaradi ilustrativnosti so prikazani rezultati le za manjši del tega signala. Opazno je boljše obnašanje (manjša napaka pri simulaciji) drugega modela, saj prvi manj ustrezno modelira šum na izhodu.

Obnašanje modelov smo ocenili še s cenilkama:

- srednja kvadratna napaka $SE = \frac{1}{N} \sum_{i=1}^N e_i^2$ in



Slika 2: Rezultati simulacije GP modela z vključenim predznanjem o šumu



Slika 3: Primerjava napak simulacij obeh GP modelov

- logaritem gostote napake [3]

$$LD = \frac{1}{2N} \sum_{i=1}^N \left(\log(2\pi) + \log(\sigma_i^2) + \frac{e_i^2}{\sigma_i^2} \right)$$

kjer je $e_i = \hat{y}_i - y_i$ napaka modela in σ_i^2 ocenjena varianca modela v i -tem koraku. Tudi vrednosti cenilk, zbrane v Tabeli 1, potrjujejo boljše obnašanje drugega modela.

4 Zaključek

Predstavljen je bil GP model za identifikacijo nelinearnih dinamičnih sistemov. Glavne prednosti GP modela pred sorodnimi metodami za identifikacijo so mera zaupanja v napoved, odvisna od učnih podatkov, majhno število parametrov, ki jih je potrebno optimizirati in možnost vključevanja različnih vrst predznanja. Prikazano je bilo, kako vključiti pred-

Tabela 1: Vrednosti cenilk za rezultata simulacije GP modelov z in brez vključenega predznanja o šumu

cenilka	“običajen”	“vključeno predznanje”
SE	0.0106	0.0045
LD	-0.88	-1.03

znanje o šumnem procesu na izhodu nekega sistema, metoda pa je bila ilustrirana na identifikaciji preprostelega dinamičnega sistema. Pri tem so se pokazale prednosti uporabe GP modela z vključenim predznanjem o šumu pred običajnim GP modelom. V prispevku smo predpostavili predznanje o modelu šuma. V kolikor tega znanja ni, lahko skušamo model šuma pridobiti iterativno z optimizacijo celotnega modela procesa kot je običajno pri parametričnih modelih.

Literatura

- [1] K. Ažman, *Identifikacija dinamičnih sistemov z Gaussovimi procesi z vključenimi lokalnimi modeli*, Magistrsko delo, Univerza v Ljubljani, 2004.
- [2] R. Murray-Smith and A. Gurrard. Gaussian Process Priors with ARMA noise models. In *Proceedings: Irish Signals and Systems Conference*, 147-153, Maynooth, 2001.
- [3] A. Girard, *Approximate Methods for Propagation of Uncertainty with Gaussian Process Models*, Ph.D. Thesis, University of Glasgow, Glasgow, 2004.
- [4] J. Kocijan, A. Girard, B. Banko and R. Murray-Smith. Dynamic Systems Identification with Gaussian Processes. *Mathematical and Computer Modelling of Dynamic Systems*, vol. 11, 411-424, 2005.
- [5] D. Matko, *Identifikacije*, Založba FE in FRI, 1998.
- [6] K.S. Narendra and K. Parthasarathy. Identification and Control of Dynamical Systems Using Neural Networks, *IEEE Transactions on NN*, Vol.1 No. 1, 4-27, 1990.
- [7] R.M. Neal, *Bayesian learning for neural networks*, Springer Verlag, New York, 1996.
- [8] C.E. Rasmussen, *Evaluation of Gaussian Processes and Other Methods for NonLinear Regression*, PhD thesis, University of Toronto, Toronto, 1996.
- [9] C.E. Rasmussen and C.K.I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge, 2006.
- [10] C.K.I. Williams. Prediction with Gaussian processes: from linear regression and beyond. In M.I. Jordan (Ed) *Learning and Inference in Graphical models*, Kluwer Academic Press, 599-621, 1998.