

Identifikacija dinamičnega sistema s histerezo z modelom na osnovi Gaussovih procesov

Kristjan Ažman[†], Juš Kocijan^{†,‡}

[†]Inštitut Jožef Stefan, Ljubljana

[‡]Politehnika Nova Gorica, Nova Gorica

kristjan.azman@ijs.si

Identification of the dynamic system with hysteresis using Gaussian process model

Dynamic system identification with Gaussian process prior model is an emerging, probabilistic, non-parametric modelling method for identification. It is comparable to artificial neural networks or fuzzy models. First advantage over these two and related methods is that the variance, associated with the model's response, is readily obtained. Another advantage is the possibility to include various prior knowledge into the model. An example of such prior knowledge is known hysteresis of the system. In this paper the incorporation of the prior knowledge of hysteresis in the GP model will be presented together with illustrative example.

1 Uvod

Za identifikacijo linearnih dinamičnih sistemov obstaja veliko metod (glej npr. [1, 2]), identifikacija nelinearnih dinamičnih sistemov pa je v splošnem težja. V članku bo predstavljena identifikacija dinamičnih sistemov z modelom na osnovi Gaussovih procesov (GP modelom), ki spada med verjetnostne neparometrične modele in je primerljiv z umetnimi nevronskimi mrežami (UNM) ali Takagi-Sugeno (T-S) mehкими modeli. 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 pomembna prednost GP modela pred UNM ali T-S mehкими modeli tudi veliko manjše število parametrov, ki jih je potrebno optimizirati. Še ena dobra lastnost GP modela je tudi možnost vključevanja različnih vrst predznanja kot npr. vključevanje lokalnih modelov [3, 4, 5], statične karakteristike, predznanja o šumu, histereze. Prav vključevanje znanja o histerezi sistema v GP model bo tema tega članka.

GP model je bil najprej uporabljen za regresijo [6], popularnost v krogu ljudi, ki se ukvarjajo s strojnimi

učanjem, pa je pridobil pozneje z deli Rasmussen-a [7] in Neal-a [8].

Evropski projekt MAC (2000-04) je spodbudil raziskovanje uporabe GP modela za identifikacijo dinamičnih sistemov, kar je botrovalo k objavi prvih prispevkov s tega področja [9, 10, 11]. 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 članku bo predstavljena uporaba GP modela za identifikacijo dinamičnega sistema s histerezo. 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 histereze v GP model. Temu bo sledil primer identifikacije nelinearnega dinamičnega sistema z GP modelom z in brez vključevanja histereze. V zaključku bo na kratko povzeto bistvo tega prispevka.

2 Modeliranje z modelom na osnovi Gaussovih procesov

Identifikacija z metodo na osnovi Gaussovih procesov je statistična, neparometrična metoda identifikacije. Bolj podrobno je predstavljena npr. v [12, 13, 7].

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: $f(\mathbf{x}_1), \dots, f(\mathbf{x}_n) \sim \mathcal{N}(0, \Sigma)$. Elementi Σ_{ij} kovariančne matrike Σ so kovariance med vrednostmi spremenljivk $f(\mathbf{x}_i)$ in $f(\mathbf{x}_j)$ ter so funkcije ustreznih 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 Σ . Ponavadi izberemo Gaussovo kovariančno funkcijo, pri čemer predpostavimo stacionarnost in gladkost procesa [14, 13]:

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

kjer je D velikost vhodnega vektorja \mathbf{x} in je $\Theta = [w_1, \dots, w_D, v]^T$ vektor parametrov, ki so imenovani hiperparametri, ker določajo verjetnostno porazdelitev funkcij in ne parametrizirajo njih samih. Hiperparameter v določa velikost kovariance in hiperparameter w_i 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 $\epsilon(t)$ beli šum z varianco v_0 , $\epsilon \sim \mathcal{N}(0, v_0)$. Predpostavimo, da je družina funkcij $f(\cdot)$ porazdeljena normalno s kovariančno funkcijo (1). Iz tega verjetnostnega okvira sledi: $y_1, \dots, y_N \sim \mathcal{N}(0, \mathbf{K})$ 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:

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

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 $\mathbf{y}|\mathbf{X} \sim \mathcal{N}(0, \mathbf{K})$, 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 dobimo z metodo največje podobnosti (ang. *maximum likelihood*), ki poišče največjo vrednost logaritma¹ porazdelitve danih vhodno-izhodnih podatkov:

$$\begin{aligned} \mathcal{L}(\Theta) &= \log(p(\mathbf{y}|\mathbf{X})) \\ &= -\frac{1}{2} \log(|\mathbf{K}|) - \frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} - \frac{N}{2} \log(2\pi) \end{aligned} \quad (5)$$

kjer je $\Theta = [w_1 \dots w_D, v, v_0]^T$ za varianco šuma v_0 razširjen vektor hiperparametrov in \mathbf{K} kovariančna

¹logaritem je uporabljen, da lahko uporabimo optimizacijo brez omejitev

matrika učnih podatkov velikosti $N \times N$. Optimizacija zahteva računanje odvoda verjetnostne porazdelitve $\mathcal{L}(\Theta)$ po vsakem od hiperparametrov, kar zahteva invertiranje kovariančne matrike \mathbf{K} v vsakem koraku optimizacije in je računsko zahtevno za velike N . Alternativna metoda za optimizacijo parametrov je, da z uporabo *a-priori* znanja izračunamo porazdelitev z numeričnim integriranjem [7].

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

$$p(y^*|\mathbf{y}, \mathbf{X}, \mathbf{x}^*) = \frac{p(\mathbf{y}, y^*)}{p(\mathbf{y}|\mathbf{X})} \quad (6)$$

Ta porazdelitev je Gaussova [13] s srednjo vrednostjo in varianco:

$$\mu(\mathbf{x}^*) = \mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1} \mathbf{y} \quad (7)$$

$$\sigma^2(\mathbf{x}^*) = k(\mathbf{x}^*) - \mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}^*) + v_0 \quad (8)$$

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. Če nov vhod leži daleč stran od učnih vhodov bo izraz $\mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}^*)$ iz (8) majhen in napovedana varianca $\sigma^2(\mathbf{x}^*)$ bo velika.

Področja vhodnega prostora, kjer je odvisnost podatkov kompleksna ali kjer je na voljo malo podatkov ali so le-ti 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:

$$\mathbf{x}_R(k) = [y(k-1), \dots, y(k-L), u(k-1), \dots, u(k-L)]^T \quad (9)$$

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 [15, 9, 5].

2.2 Vključevanje predznanja o histerezi v GP model

Histereza je lastnost sistema, da sistem ne sledi takoj vzbujanju, ampak reagira počasi oziroma se ne vrne v originalno stanje, t.j. izhod sistema (2) ni odvisen samo od njegove (bližnje) zgodovine $\mathbf{x}_R(k)$, ampak tudi od stanja $\eta(k)$, v katerem se sistem trenutno nahaja [16].

$$y(k) = f(\mathbf{x}_R(k), \eta(k)) + \epsilon(k) \quad (10)$$

Ena izmed možnosti, kako v GP model vključiti znanje o histerezi, je razširitev vhodnega vektorja (9) še za en element, ki predstavlja stanje sistema η . Razširjeni vektor regresorjev $\mathbf{x}(k)$ je tako:

$$\mathbf{x}(k) = [\mathbf{x}_R(k), \eta(k)]^T \quad (11)$$

Pri gradnji GP modela vzorce s statične karakteristike sistema, ki sedaj vključujejo tudi podatek o stanju histereze, vstavimo kot del učnih točk v GP model.

3 Primer modeliranja histereze z GP modelom

3.1 Primer dinamičnega sistema s histerezo

Kot primer za prikaz prednosti, ki jih prinaša vključevanje predznanja v GP model, smo si izbrali nelinearni dinamični sistem, ki je bil dobljen tako, da je bila v sistem (12), uporabljen že v [17]:

$$y(k+1) = \frac{y(k)}{1+y^2(k)} + u^3(k) \quad (12)$$

vključena histereza na izhodu z višino $\Delta y_H = \pm 0.5$ in širino $-1.4 < y < 1.4$. Za izhod iz sistema velja:

$$y_H(k+1) = y(k+1) + \Delta y_H(\eta(k)) + \epsilon(k) \quad (13)$$

kjer se v vsakem koraku k ob določitvi izhoda $y(k+1)$ posodobi tudi stanje sistema $\eta(k+1)$:

$$\eta(k+1) = \begin{cases} 1, & \text{ko } y(k) \geq l_h = 1.4 \\ -1, & \text{ko } y(k) \leq l_l = -1.4 \\ \eta(k), & \text{drugače.} \end{cases} \quad (14)$$

Izhod sistema $y_H(k+1)$ je bil moten z belim šumom z varianco $v_0 = 4 \cdot 10^{-4}$.

Naš cilj je bila izgradnja GP modela, ki bo opisoval dinamično obnašanje sistema (13) na področju, omejenem z vhomom u med $u_{\min} = -1.5$ 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 histerezi in
- GP model z vključenim predznanjem o histerezi.

3.2 Učenje GP modelov

Za učenje smo uporabili vzorce iz statične karakteristike sistema in vzorce, ki smo jih dobili iz dinamičnega odziva modela. Za vzbujanje sistema, kjer nas je zanimala dinamika, je bil uporabljen psevdonaključen binarni signal (PRBS) [2] s taktom $k_{sw} =$

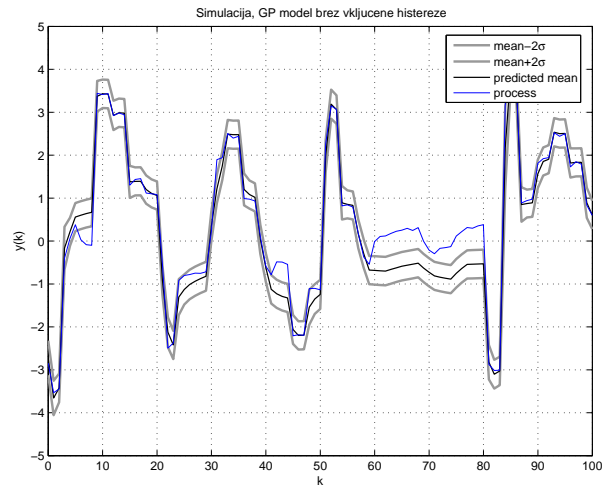
3, kjer pa amplituda ob spremembi lahko zavzame katerokoli vrednost med u_{\min} in u_{\max} .

Najprej je bil narejen GP model za sistem (13) brez vključevanja znanja o histerezi v model. Iz statične karakteristike in dinamičnega odziva modela smo z vzorčenjem dobili 242 vektorjev, ki so opisovali obnašanje sistema in so bili uporabljeni za učenje. Ob vzorčenju v k -tem koraku sta vhodni del učnega vektorja sestavljala vhod in izhod iz sistema $[u(k-1), y(k-1)]$, izhodni del učnega vektorja pa izhod $y(k)$.

Za učenje drugega modela, t.j. GP modela z vključenim znanjem o histerezi, so bili uporabljeni isti vzorci kot za učenje prvega modela, le da so imeli vhodni deli učnih vzorcev eno komponento več – podatek o stanju sistema $\eta(k)$.

3.3 Rezultati

Za ovrednotenje obeh modelov smo modela simulirali z drugim PRBS signalom, kot je bil uporabljen za učenje in primerjali rezultate simulacije s pravim odzivom modela. Zaradi ilustrativosti so prikazani rezultati le za manjši del tega signala. Rezultate ovrednotenja prvega modela vidimo na Sliki 1, rezultate ovrednotenja drugega pa na Sliki 2.



Slika 1: Simulacija običajnega GP modela

Opazno je boljše obnašanje drugega modela, saj ima prvi model nekaj težav na področju, kjer je sistem histerezo obnašanje, t.j. $l_l \leq y(k) \leq l_h$.

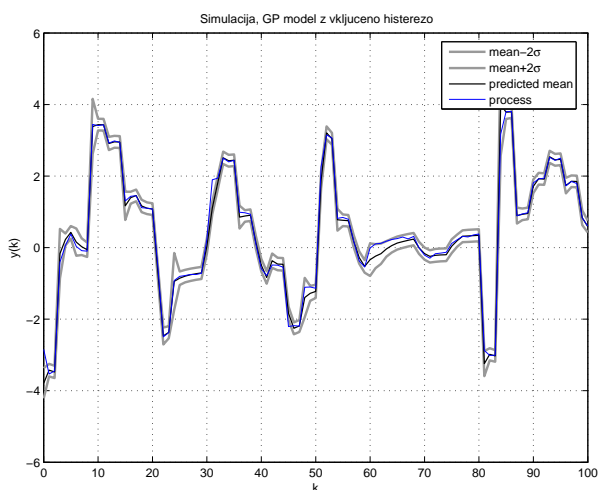
Obnašanje modelov smo ocenili še s cenilkama:

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

- logaritem gostote napake [15]

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

kjer je $e_i = \hat{y}_i - y_i$ napaka modela in σ_i^2 ocenjena varianca modela v i -tem koraku. Tudi vrednosti ce-



Slika 2: Simulacija GP modela, v katerega je bilo vključeno znanje o histerezi

nilk, zbrane v Tabeli 1, potrjujejo boljše obnašanje drugega modela.

Tabela 1: Vrednosti cenilk za rezultata simulacije GP modelov z in brez vključene histereze

cenilka	“brez histereze”	“s histerezo”
SE	0.19	0.035
LD	2.4	-0.41

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 predznanje o histerezi nekega sistema, metoda pa je bila ilustrirana na identifikaciji preprostega dinamičnega sistema. Pri tem so se pokazale prednosti uporabe GP modela z vključenim predznanjem o histerezi pred običajnim GP modelom.

Literatura

[1] L. Ljung, *System Identification – Theory for the User*, 2nd Edn., Prentice Hall, New Jersey, 1999.

[2] D. Matko, *Identifikacije*, Založba FE in FRI, 1998.

[3] E. Solak, R. Murray-Smith, W.E. Leithead, D.J. Leith and C.E. Rasmussen. Derivative observations in Gaussian Process Models of Dynamic Systems, In S.

Becker, S. Thrun and K. Obermayer (Eds) *Advances in Neural Information Processing Systems 15*, MIT Press, 529–536, 2003.

- [4] J. Kocijan and D.J. Leith. Derivative Observations used in Predictive Control. In *Proceedings: Melecon 2004*, Dubrovnik, 2004.
- [5] K. Ažman, *Identifikacija dinamičnih sistemov z Gaussovimi procesi z vključenimi lokalnimi modeli*, Magistrsko delo, Univerza v Ljubljani, 2004.
- [6] A. O’Hagan. On curve fitting and optimal design for regression (with discussion). *Journal of the Royal Statistical Society B*, vol. 40, 1-42, 1978.
- [7] C.E. Rasmussen, *Evaluation of Gaussian Processes and Other Methods for NonLinear Regression*, PhD thesis, University of Toronto, Toronto, 1996.
- [8] R.M. Neal, *Bayesian learning for neural networks*, Springer Verlag, New York, 1996.
- [9] J. Kocijan, A. Girard, B. Banko and R. Murray-Smith. Dynamic Systems Identification with Gaussian Processes. In *4th Mathmod, Wien, 2003*, razširjena verzija v *Mathematical and Computer Modelling of Dynamic Systems* (v tisku).
- [10] J. Kocijan, B. Banko, B. Likar, A. Girard, R. Murray-Smith and C.E. Rasmussen. A case based comparison of identification with neural networks and Gaussian process models. In *Proceedings: IFAC Intelligent Control Systems and Signal Processing Conference*, Faro, 137–142, 2003.
- [11] G. Gregorčič and G. Lightbody. Gaussian Processes for Modelling of Dynamic Non-linear Systems. In *Proceedings: Irish Signals and Systems Conference*, Cork, 2002.
- [12] 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.
- [13] M.N. Gibbs, *Bayesian Gaussian Processes for Regression and Classification*, Ph.D. Thesis, Cambridge University, Cambridge, 1997.
- [14] D.J. Mackay. Gaussian Processes. *Information theory, inference and learning algorithms*, Cambridge University Press, 2003.
- [15] A. Girard, *Approximate Methods for Propagation of Uncertainty with Gaussian Process Models*, Ph.D. Thesis, University of Glasgow, Glasgow, 2004.
- [16] Wikipedia, the free encyclopedia, *Hysteresis* povezava: <http://en.wikipedia.org/wiki/Hysteresis>, datum objave: 7.7.2005
- [17] 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.