



KOOPERÁCIÓ ÉS GÉPI TANULÁS LABORATÓRIUM

Kernel módszerek – idősor előrejelzés Mérési útmutató

Készítette:
Engedy István
(engedy@mit.bme.hu)

Méréstechnika és Információs Rendszerek Tanszék
Budapesti Műszaki és Gazdaságtudományi Egyetem

2009. szeptember 1.

Tartalomjegyzék

1. fejezet - Felkészülési feladatok.....	3
2. fejezet - A kiadott kód és a toolboxok leírása	4
3.1. NNSYSID toolbox, lipschit függvény	4
3.2. SVM-KM toolbox, függvényapproximációban használatos függvények	4
3.3. mgts.m.....	4
3.4. maketstp.m.....	5
3. fejezet - Mérési feladatok	6
1. Feladat – modell-fokszám meghatározás.....	6
2. Feladat – idősor-előrejelzés SVM-mel.....	6
3. Feladat – idősor előrejelzés MLP–vel	6
4. fejezet - Jegyzőkönyv.....	6
5. fejezet - Leadás.....	6
Irodalomjegyzék	7

1. fejezet - Felkészülési feladatok

Rendszermodellezési feladatoknál az első feladat az apriori információk szerzése után a modellstruktúra választása. Előrejelzési feladatoknál ez általában az NFIR vagy NARX struktúra. Az ilyen modellek fokszámának megállapítására több lehetőség kínálkozik. Az egyik ilyen lehetőség a keresztkiértékelés alkalmazása, amikor különböző fokszámú modelleket konstruálunk és ezek közül a validációs eredmények alapján a legjobbkat kiválasztjuk. Itt fontos szerepet töltenek be az ún. információs kritériumok. Egy másik lehetőség a Lipschitz-indexen alapuló heurisztikus eljárás, ami a fokszám(ok) mintapontokból történő közvetlen becslésére alkalmas.

Bővebben: Horváth Gábor et al: Neurális hálózatok (pp 252-256)

A Lipschitz-index meghatározása az NNSYSID nevű MATLAB toolbox használatával könnyen megoldható.

A Mackey-Glass kaotikus folyamatot gyakran alkalmazzák modell-fokszám becslési eljárások tesztelésére. A Mackey-Glass nemlineáris differenciálegyenlettel (többek között) az emberi testben a fehérvérsejtek termelődésének dinamikáját modellezzik. Mivel a folyamatban egy késleltetés is megjelenik periodikus jellegű kaotikus folyamatot kapunk. A $\tau < 17$ értékekre a kapott idősor periodikus. A folyamat differenciálegyenlete a következő:

$$\frac{dx(t)}{dt} = -0.1 x(t) + \frac{0.2 x(t - \tau)}{1 + x^{10}(t - \tau)}$$

ahol τ egy paraméter. A folyamatot a t változó egész értékei szerint mintavételezve a következő rekurzív összefüggést kaphatjuk:

$$x[t] = 0.9 x[t - 1] + \frac{0.2 x[t - \tau]}{1 + x^{10}[t - \tau]}$$

Gyakori feladat a rendszermodellezési problémák között a jóslás, előrejelzés. Például valutaárfolyamok, tőzsde, időjárás, stb. Ennek során a rendszer múltbeli viselkedésének modellezésével történik a jövőbeli viselkedés előrejelzése. Ez gyakran egy idősor alakulásának előrejelzését jelenti.

Az idősor-előrejelzési feladatok gyakran legjobb megoldási módszerei neurális hálózatok segítségével történnek, noha bonyolult statisztikai módszerekkel is megoldhatóak.

Bővebben: Horváth Gábor et al: Neurális hálózatok (pp 263-267)

2. fejezet - A kiadott kód és a toolboxok leírása

3.1. NNSYSID toolbox, lipschit függvény

Az NNSYSID toolbox nemlineáris dinamikus rendszerek neurális hálózat alapú identifikációjához szükséges eszközöket tartalmaz. A Lipschitz-index kiszámítására a *lipschit* függvény szolgál. Használatához lásd alább a függvény leírását. Mivel a függvény a különböző modell-paraméterekre adja meg a Lipschitz-index értékét, ezért ezek közül nekünk kell kiválasztani az optimálisat. Ez a függvény meredek letörésének végénél található. Általában egy ilyen meredek letörés után az érték közel állandó lesz. Az ehhez az értékhez tartozó fokszám a keresett paraméter.

Tipp: A Lipschitz-index kiszámításakor – feladattól függően – feltételezhető, hogy a keresett paraméter bizonyos korlátok között lesz, akkor érdemes ebben a tartományban vizsgálni a kapott értékeket, így az 1-es fokszámhoz tartozó esetlegesen nagyon nagy érték nem befolyásolja a fokszám meghatározását, másrészt a futásidő nagyban csökkenthető.

```
[OrderIndexMat]=lipschit(U,Y,m,n)

INPUTS:
  U - Sequence of inputs (row vector)
  Y - Sequence of outputs (row vector)
  m - Vector specifying the input lag spaces to investigate
  n - Vector specifying the ouput lag spaces to investigate

OUPUTS:
  OrderIndexMat - A matrix containing the order indices for each
  combination of elements in the vectors m and n. The number of rows
  corresponds to the number of elements in m, while the number of
  columns corresponds to the number of elements in n

Time series:
  U=[]; m=0;
```

Bővebben: <http://www.iau.dtu.dk/research/control/nnsysid.html>, MATLAB: *help lipschit*

3.2. SVM-KM toolbox, függvényapproximációban használatos függvények

MATLAB-hoz számos SVM toolbox létezik. Az egyik ilyen az SVM-KM nevű toolbox, melynek használata viszonylag egyszerű. Regressziós SVM létrehozása és tanítása a *svmreg* függvénnyel történik, használata pedig az *svmval* függvénnyel.

```
példa svmreg használatára:
  [xsup,ysup,w,b] = svmreg(tpx,tpy,1,0,'gaussian',0.7);

példa svmval használatára:
  y = svmval(x,xsup,w,b,'gaussian',0.7);
```

Bővebben: <http://asi.insa-rouen.fr/enseignants/~arakotom/toolbox/index.html>, MATLAB: *help svmreg, help svmval*

3.3. mgts.m

Ez a függvény szolgál a Mackey-Glass idősor generálására. Használatához meg kell adni a generálni kívánt idősor hosszát, illetve egy tau paramétert. Erről bővebben lásd a bevezetőt.

```
data = mgts(n,tau);
```

Inputs:

```
  n    -   number of samples
  tau  -   the tau constant in the equation
```

Output:

```
  data -   the time series vector (row)
```

Example:

```
  data = mgts(500,12);
```

3.4. maketstp.m

Ez a függvény arra szolgál, hogy idősorokból egyszerű statikus tanuló eljárásokhoz olyan tanítómintákat generáljon, melyben a bemenet az idősor egy n hosszú egybefüggő szakasza, a kimenete pedig az idősor ezen szakaszát követő elem. Értelem szerűen a függvény bemenetei az idősor, mely egy vektor, illetve n , a tanítóminta bemenetének hossza, kimenete pedig X , a tanító minták bemenetét tartalmazó $n*k$ mátrix, illetve Y , a tanító minták kimenetei, mely egy vektor. A függvény által generált tanítópontokkal egy statikus tanuló eljárás úgy használható, mintha NFIR struktúrájú lenne.

```
[X,Y] = maketstp(ts,n);
```

Inputs:

```
  ts    -   the time series vector
  n     -   the length of the training sample input (X(k))
```

Outputs:

```
  X     -   the training sample inputs (n*k matrix)
  Y     -   the training sample targets (k row-vector)
```

Example:

```
  ts = [1:1000];
  [tpx,tpy] = maketstp(ts,5);
```

3. fejezet - Mérési feladatok

1. Feladat – modell-fokszám meghatározás

Határozza meg a Mackey-Glass folyamat modell-fokszámát különböző τ paraméterekre (pl: 5, 10, 15, 20, stb) a Lipschitz-index kiszámításával, majd vesse össze azt a τ paraméter értékével!

Vizsgálja meg a $\tau = 17$ paraméterű, 1:6 arányban alulmintavételezett Mackey-Glass folyamat Lipschitz-indexének alakulását különböző amplitúdójú Gauss-zaj hozzáadására!

2. Feladat – idősor-előrejelzés SVM-mel

Készítsen NFIR struktúrájú SVM-et a Mackey-Glass folyamat előrejelzésére. Tanítómintának használja a folyamat egy megfelelően hosszú idősorát. Vizsgálja meg a rendszer átlagos négyzetes hibáját mind a tanító idősoron, mind a tanítóminta utáni szakaszon. A bemeneti értékeket az eredeti idősorból vegye.

Végezze el a mérést különböző modell-fokszámokkal (NFIR fokszáma), és vesse össze az eredményt az adott paraméterű Mackey-Glass folyamat Lipschitz-indexével. Vizsgáljon meg több különböző τ paraméterű folyamatot (pl: 5, 10, 15, 20, stb).

3. Feladat – idősor előrejelzés MLP-vel

Készítsen NFIR struktúrájú MLP hálót, a Mackey-Glass folyamat előrejelzésére ($\tau = 17$). Használja fel a már előrejelzett értékeket további előrejelzések kiszámításához. Próbálja minimalizálni a tanító idősor hosszát.

4. fejezet - Jegyzőkönyv

A jegyzőkönyvnek – a megoldók nevéen és neptun-kódján kívül – a következőket kell tartalmaznia minden egyes részfeladatra vonatkozóan:

- KI-MIT oldott meg (feladat felosztása és értelmezése).
- HOGYAN oldották meg (megoldás leírása, nemtriviális algoritmusok).
- MIÉRT így oldották meg (megoldás indoklása).
- EREDMÉNYEK összefoglalása (értelmezés és értékelés).
- A jegyzőkönyv mellett a megfelelően KOMMENTEZETT KÓD.

5. fejezet - Leadás

A megírt matlab forrásokat és a jegyzőkönyvet együtt, tömörített formában (egy MSCLAB1-7_NEPTUN1_NEPTUN2.ZIP-ben) kell a megadott határidőig (laborgyakorlati héten péntek éjfélig) e-mailben – fájlként csatolva – az engedy@mit.bme.hu címre elküldeni.

A levél subject mezéjében az „[MScLab1-7]” szöveg szerepeljen, a levél szövege pedig rendre a megoldók neve, és neptun-kódja legyen. Több megoldó esetén az email-t a többi megoldónak is CC-zni kell!

Irodalomjegyzék

- [1] Altrichter M., Horváth G., Pataki B., Strausz Gy., Takács G., Vallyon J., *Neurális Hálózatok*, Panem, 2006
- [3] Magnus Nørgaard, "The NNSYSID Toolbox",
<http://www.iau.dtu.dk/research/control/nnsysid.html>
- [3] S. Canu and Y. Grandvalet and V. Guigue and A. Rakotomamonjy, "SVM and Kernel Methods Matlab Toolbox" Perception Systèmes et Information, INSA de Rouen, Rouen, France, 2005,
<http://asi.insa-rouen.fr/enseignants/~arakotom/toolbox/index.html>