

Neuronové časové řady (ANN-TS)

Menu:	QCExpert	Prediktivní metody	Neuronové časové řady
-------	----------	--------------------	-----------------------

Tento modul (*Artificial Neural Network Time Series* – ANN-TS) využívá modelovacího potenciálu neuronové sítě k predikci a předpovědi budoucích hodnot jednorozměrné časové řady. Za jednorozměrnou časovou řadu zde považujeme posloupnost n naměřených hodnot y_i , o nichž se předpokládá, že

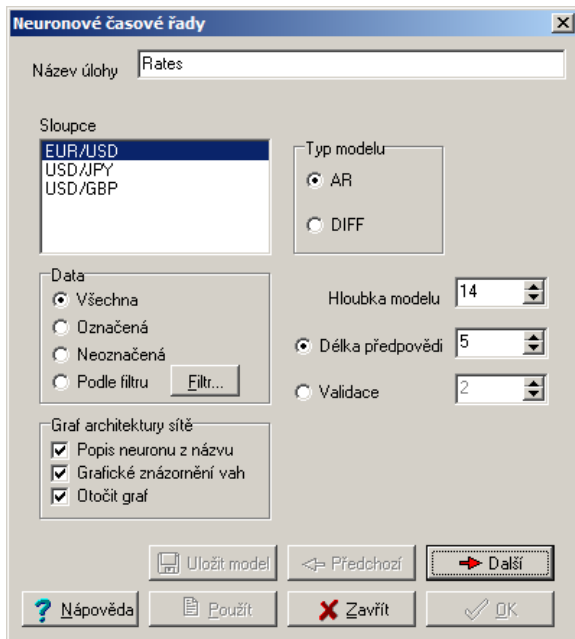
- jsou naměřeny v pravidelných časových okamžicích (takže informaci o čase měření lze nahradit pouhým pořadovým číslem měření bez ztráty informace), případně se jedná o posloupnost jednotlivých diskretních entit (např. jednotlivé výrobky vyráběné po sobě – ne nutně v pravidelných časových intervalech);
- mohou nabývat libovolných reálných hodnot (to znamená, že nejsou omezeny jen na celá čísla, případně nečíselné hodnoty, tedy případná předpověď reálných čísel bude dávat smysl);
- každá naměřená hodnota nějak závisí na hodnotách naměřených před ní, tedy
$$y_i = G(y_{i-1}, y_{i-2}, \dots, y_{i-r}) + \varepsilon_i, \text{ zkráceně } G(i, r) + \varepsilon_i, \text{ kde } r < n \text{ nazveme hloubkou modelu.}$$

Jako příklady takových časových řad lze uvést technologické veličiny, periodicky měřené výsledky výrobních procesů, nestabilita procesů, parametry přírodních systémů, vývoj finančních a ekonomických indexů, kurzů, cen a podobně.

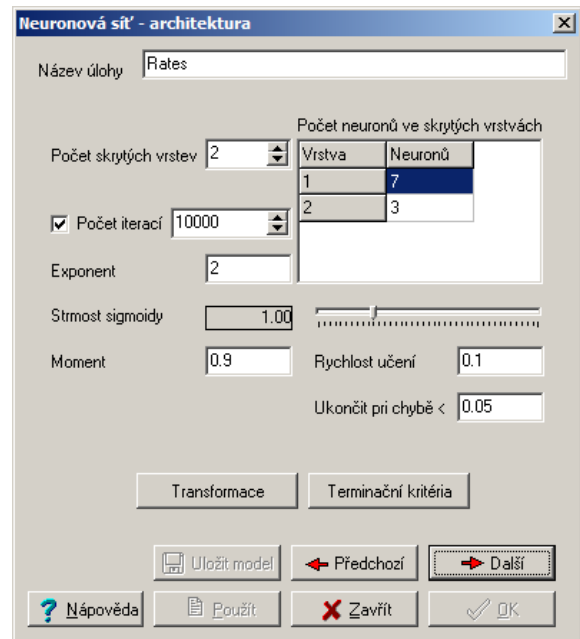
Nalezením vztahu $G()$ lze pak predikovat z r -tice posloupnosti naměřených hodnot hodnotu následující. Rozlišujeme dva druhy predikce – predikci hodnoty, která byla naměřena, tedy lze odhadnout chybu predikce $y_k - G(k, r)$, a předpověď (prognózu) hodnot v budoucnosti, které ještě neznáme: $\hat{y}_{n+1} = G(n+1, r)$. Pokud přidáme předpověděnou hodnotu k hodnotám již známým a budeme ji považovat za správně odhadnutou, můžeme rekurzivně předpovídat další neznámé hodnoty v budoucnosti. Je ovšem třeba si uvědomit, že každá další předpověď je zatížena rostoucí chybou, jejíž velikost není známa. To je částečně vyřešeno v následujícím modulu ANN-TS: Para-Bootstrap, kde je možné simulovat intervaly spolehlivosti předpovědi. Jako model G použijeme neuronovou síť s r vstupními proměnnými a jednou výstupní proměnnou. Tento modul nabízí dva typy modelů: Autoregresní model vycházející přímo z naměřených hodnot (model AR) a model vycházející z prvních diferencí (model DIFF). Obecně se modely typu AR spíše hodí pro modelování stacionárních řad, kdy se budoucí hodnoty nedostanou mimo interval, v němž se nacházejí již naměřené hodnoty. Model typu DIFF je spíše vhodný tam, kde není vhodný model AR, tedy pro nestacionární řady, řady se systematickým lineárním trendem a řady, u nichž se očekávají v budoucnosti hodnoty mimo dosavadní minimum a maximum. Další podrobnosti a popis je shodný s kapitolou neuronové sítě.

Data a parametry

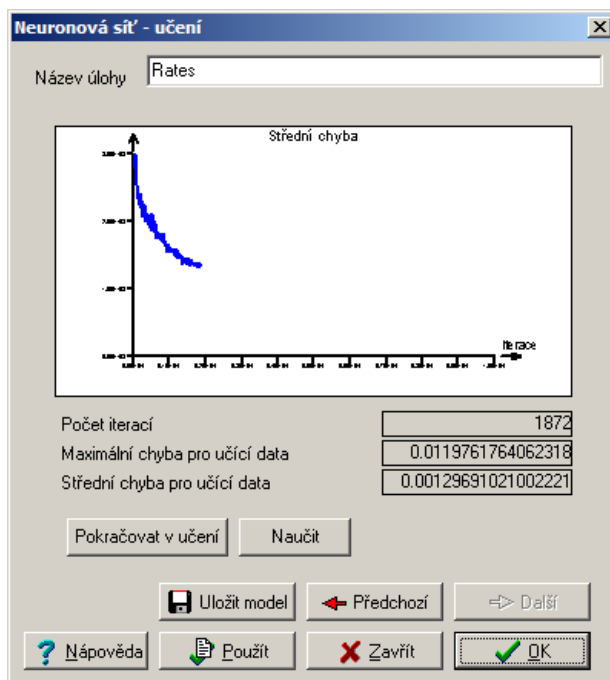
Data pro tento model tvoří jeden sloupec v datovém listu. Detaily výpočtu se nastavují v sérii 3 dialogových oken (viz Obrázek 1, Obrázek 2, Obrázek 3). V prvním okně se vybere sloupec s daty, typ modelu (AR nebo DIFF), hloubka modelu (počet hodnot z nichž se bude predikovat následující hodnota), délka předpovědi, případně délka validace modelu.



Obrázek 1 Výběr proměnných



Obrázek 2 Topologie sítě a parametry optimalizace



Obrázek 3 Okno řízení optimalizace sítě

Délka předpovědi definuje počet předpovědí budoucích, dosud neznámých hodnot. U většiny reálných řad lze předpovědět jen několik hodnot do budoucnosti. U řad s výraznou (i když složitou) periodickou složkou lze úspěšně předpovídat i desítky hodnot. Validace slouží k orientační představě o přesnosti a úspěšnosti předpovědi. Provede se předpověď posledních q naměřených hodnot z hodnot předcházejících. Počet q validovaných hodnot zadá uživatel v políčku *Validace*. Nelze současně validovat a předpovídat, proto je vhodné provést nejdříve výpočet s validací a po posouzení úspěšnosti teprve provést předpověď odpovídajícího počtu hodnot v budoucnosti.

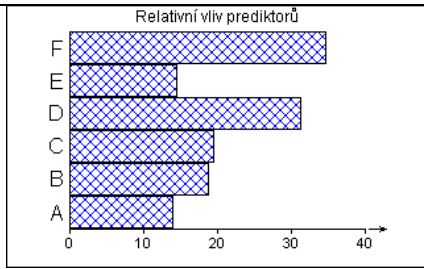
Význam a popis dalších oken a funkcí je podobný jako v kapitole Neuronové sítě. Tlačítkem *Uložit model* se aktuálně vypočítaný model uloží do souboru pro pozdější použití v modulu *Predikce*, případně k automatickým predikcím v inteligentní databázi QCE-DataCenter®.

Protokol

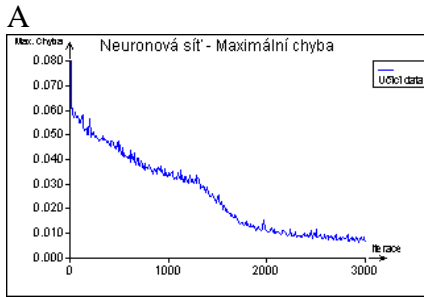
Název úlohy	Zadaný název úlohy
Data	
Typ modelu	Zvolený typ modelu – AR nebo DIFF
Hloubka modelu	Zvolený počet prediktorů pro předpověď
Délka předpovědi	Zvolená délka předpovědi
Validace	Zvolený počet dat na konci řady použitých pro validaci modelu
Nezávisle proměnné	Seznam nezávisle proměnných jako číslovaný seznam zvoleného sloupce s časovým posunem Tedy pro sloupec X bude X3 znamenat X[i-3]
Typ transformace	Typ transformace nezávisle proměnných
Závisle proměnné	Název vybraného sloupce
Typ transformace	Typ transformace závisle proměnných
Vrstva, Neuronů	Číslo vrstvy a počet neuronů ve vrstvě
Strmost sigmoidy	Zadaná strmost sigmoidy
Moment	Zadaný moment
Rychlost učení	Zadaná rychlost učení
Ukončit při chybě	Kritérium chyby pro ukončení
Procent dat pro učení (%)	Procent dat pro učení (%), je-li vybráno
Podmínky ukončení optimalizace	Zadané terminační podmínky pro ukončení výpočtu
Výpočet	Informace o průběhu výpočtu
Počet iterací	Skutečný počet iterací
Maximální chyba pro učící data	Dosažená maximální chyba pro učící data
Střední chyba pro učící data	Dosažená střední chyba pro učící data
Maximální chyba pro testovací data	Dosažená maximální chyba pro testovací data
Střední chyba pro testovací data	Dosažená střední chyba pro testovací data
Váhy	Vypočítané optimální váhy neuronové sítě
Vrstva / Neuron	Číslo vrstvy a číslo neuronu
Relativní vliv	Relativní vliv X[i-j] na X[i] určený podle normy příslušných vah v neuronové síti.
Časová řada	Tabulka naměřených hodnot a jejich predikcí a reziduí.

Grafy

	<p>Graf časové řady s predikcí nebo validací (dle zadané volby). Zobrazuje naměřené hodnoty (modře) a vypočítané hodnoty (zeleně). Svislá příčka označuje začátek předpovědi, nebo validace. Od tohoto místa vpravo nemá model k dispozici data a zelené hodnoty předpovědi jsou vypočítané pouze na základě dat vlevo od této svislé hranice.</p>
	<p>Graf reziduí, tedy odchylek naměřených hodnot od predikce, nebo předpověděných hodnot od validačních dat.</p>
	<p>Grafické vyjádření architektury sítě. Byla-li při výpočtu vybrána možnost <i>Grafické znázornění vah</i>, pak tloušťka synapsí znázorňuje absolutní velikost odpovídající váhy a tím také množství informace, která teče mezi dvěma neurony. Podle tloušťky synapse se tak dá usuzovat na významnost nezávisle proměnné (vychází z ní tlusté synapse) a také na kvalitu predikce jednotlivých závisle proměnných (vchází do ní tlusté synapse). Barva synapsí určuje pouze znaménko váhy (červená = záporná váha, modrá = kladná váha). Byla-li vybrána možnost <i>Popis neuronu z názvu</i>, použije se název sloupce jako popis vstupních a výstupních neuronů.</p>



Relativní vliv prediktorů na predikci vyjádřený jako součet absolutních vah jednotlivých prediktorových proměnných. Tato statistika je pouze mírou vlivu proměnných na odezvu, není testem významnosti. Tento graf má reálnou vypovídací schopnost pouze při vhodné volbě modelu.



Graf učicího procesu, pokles součtu čtverců rozdílů predikce a skutečných hodnot závisle proměnné v závislosti na počtu iterací. Na obrázku je typický úspěšný učicí proces, který postupně zlepšoval model pro zadaná data.