

Prostriedky identifikácie zložitých systémov

Resources of identification of complex systems

Michaela Horalová Kalinová, Katedra aplikovanej informatiky, FPV UCM v Trnave

Stanislav Horal, Katedra aplikovanej informatiky, FPV UCM v Trnave

Abstract: Growing complexities of problems arising in different areas of real world lead to the necessity to apply a wide range of methods and tools for a deeper analysis of understanding the fundamental problem. Defining the fundamental problem helps to build such a model whose properties and behaviour would correspond to the nature of the problem. In the process of complex technical system modelling, great amounts of data of information value are supposed to be generated as a result of a wide range of experiments. Therefore, it is necessary to consider an optimal mode of data organisation. The aim of this contribution is to present ways of technical system identification using the MATLAB program, an interactive powerful program for scientific and technical calculations. We will also consider the possibilities of exploitation of data mining and data warehousing technologies in the process of identification.

Key words: data warehouse, real time

Abstrakt: Narastajúca zložitost' problémov v rozličných oblastiach reálneho sveta vedie k nutnosti aplikovania širokej škály metód a nástrojov pre hlbšiu analýzu pochopenia podstaty problému. Definovanie podstaty problému vedie k vybudovaniu takého modelu, ktorý by svojimi vlastnosťami a správaním zodpovedal jeho povahe. Keďže pri modelovaní zložitých technických systémov sa predpokladajú veľké objemy dát s informačnou hodnotou, ktoré sú výsledkami širokej škály experimentov, je potrebné uvažovať o optimálnom spôsobe ich organizácie. Cieľom príspevku je predstaviť spôsoby identifikácie technických systémov pomocou programu MATLAB, výkonného interaktívneho prostredia pre vedecko-technické výpočty. V článku sa tiež budeme zaoberať zhodnotením možností využívania technológií dolovania dát a dátových skladov v procese identifikácie.

Kľúčové slová: dátový sklad, reálny čas

1. Úvod

Pojem identifikácie systémov môžeme vo všeobecnosti predstaviť ako spôsob nájdenia vhodného matematického modelu príslušného systému pre konkrétny účel. Na proces identifikácie existujú dva základné pohľady t.j. pri identifikácií vychádzame z matematicko-fyzikálnej analýzy (analytická identifikácia), alebo za smerodajné považujeme merania na systéme (experimentálna identifikácia). V príspevku sa budeme venovať experimentálnej identifikácií vo forme „off line“ identifikácie, pričom budeme predpokladať, že merané údaje budeme ukladať do vhodných dátových úložísk reprezentovaných prostredníctvom dátového skladu a následne vyhodnocovať v prostredí MATLABu. Predmetom analýzy budú rozľahlé technologické systémy, t.j. rozľahlé systémy, ktorých vstupné a zodpovedajúce výstupné údaje zodpovedajú technologickým meraniam.

Cieľom príspevku je predstaviť súčasné automatizované možnosti identifikácie technických systémov a zhodnotiť možnosti technológií dolovania dát a dátových skladov v tejto oblasti.

2. Dátový sklad reálneho času

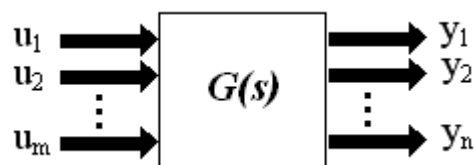
Dátový sklad (Data Warehouse) môžeme predstaviť ako predmetne orientovanú, integrovanú, časovo premenlivú kolekciu údajov primárne určenú pre proces podpory rozhodovania. Jedným z hlavných cieľov dátových skladov je integrácia údajov z rozličných, typicky nehomogénnych dátových zdrojov. Z pohľadu užívateľa je organizácia dát predmetne orientovaná, t.j. organizovaná v zmysle jednotlivých oblastí záujmu. Základnou organizačnou jednotkou dátového skladu je multidimenzionálna štruktúra. Údaje organizované do takejto štruktúry slúžia ako podklad pre získanie sumarizovaných a agregovaných údajov. Hlavnými výhodami tohto riešenia je rýchly a komplexný prístup k veľkému objemu dát, možnosť komplexnej analýzy, modelovania a predikcie. Základnú entitu multidimenzionálnej štruktúry tvorí dimenzia, ktorá obsahuje dáta z vybranej časti predmetnej oblasti. (1, 2)

Pre analýzu a riadenie technických systémov má zásadný význam spracovávanie údajov, ktoré vznikajú ako výsledky meraní technologických procesov, v reálnom čase. Požiadavka na spracovávanie údajov takéhoto charakteru môže byť realizovaná prostredníctvom dátových skladov reálneho času, ktoré dokážu v reálnom čase zabezpečovať proces ETL, t.j. fázu extrakcie (extraction), fázu transformácie (transformation) a fázu načítavania údajov do dátového skladu (loading). Z hľadiska reprezentácie výsledkov meraní technologických procesov nedochádza k nutnej zmene reprezentácie multidimenzionálnej štruktúry.

Využitie možností dátových skladov a technológie dolovania dát má význam najmä pre hierarchické systémy, kde je možné analyzovať systémy na jednotlivých úrovniach hierarchického systému ako aj vzájomné vzťahy medzi podsystémami. Výsledky analýz predstavujú znalosti, ktoré majú význam najmä pre proces rozhodovania pri riadení a optimalizácii riadenia zložitých systémov.

3. Reprezentácia experimentálnych údajov

Mnohorozmerný systém môžeme predstaviť ako konečnú množinu systémov so vzájomnými väzbami rozmiestnenými v priestore. Pri identifikácii takéhoto mnohorozmerného systému má zásadný význam analýza informačných tokov, ktoré jednoznačne definujú väzby medzi elementmi komplexného systému. Z hľadiska dynamiky mnohorozmerných systémov je potrebné správne určiť parametre systému, ktorých zmeny v čase budú predmetom skúmania a ktoré budú reprezentovať zmeny stavu systému v čase. Stav mnohorozmerného systému v čase t môžeme definovať ako množinu okamžitých hodnôt vybraných parametrov, pričom neberieme do úvahy správanie sa systému v čase $t_a < t$. V zmysle technologických systémov hodnoty parametrov predstavujú hodnoty meraných veličín nameraných v jednotlivých miestach rozľahlého systému. Na základe týchto predpokladov môžeme dynamiku mnohorozmerného systému vyjadriť ako postupnosť stavov systému v čase. Celkový stav riadeného mnohorozmerného systému reprezentujú okamžité hodnoty vstupných veličín $u(t)$, výstupných veličín riadeného systému $y(t)$ a hodnoty nameraných stavových veličín $x(t)$.



Obrázok 1: Mnohorozmerný systém

Pre reprezentáciu experimentálnych údajov prostredníctvom multidimenzionálnej štruktúry dátového skladu je optimálne reprezentovať mnohorozmerný systém v stavovom priestore. Pri návrhu vychádzame z diskrétného stavového modelu v štandardnom tvare:

$$\begin{aligned} x(k+1) &= A.x(k) + B.u(k) \\ y(k) &= C.x(k) + D.u(k) \end{aligned} \quad ,$$

kde $x(k)$ predstavuje stĺpcový vektor hodnôt stavových veličín, $u(k)$ je stĺpcový vektor hodnôt vstupných veličín a $y(k)$ je stĺpcový vektor výstupných veličín v k -tom okamžiku merania.

Pre odhalenie vzájomných vnútorných väzieb a korelácií medzi veličinami viacrozmerného systému musí každá relevantná merateľná veličina tvoriť samostatnú dimenziu multidimenzionálnej štruktúry. Časová dimenzia multidimenzionálnej štruktúry je nositeľom hodnôt reprezentujúcich diskkrétne časy merania.

4. Prostriedky automatizovanej identifikácie

Proces identifikácie môžeme klasifikovať do viacerých fáz: meranie a uchovávanie údajov, odhad štruktúry modelu, odhad parametrov analyzovaného systému. Vhodným prostriedkom pre automatizovaný proces identifikácie je výkonné interaktívne prostredie MATLAB, určené pre vedecko-výskumné výpočty. Prostredie MATLAB poskytuje kolekciu nástrojov tzv. identifikačný toolbox, pre identifikáciu systémov, ktoré je možné rozličným spôsobom dátovo reprezentovať. Možnosti identifikačného toolboxu znázorňuje nasledovný prehľad:

- **Lineárne modely identifikácie**

- identifikácia modelov s frekvenčnou charakteristikou
 - § metóda SPA (**S**Pectral Analysis – Blackman-Tukey)
 - § metóda SPAFDR (**S**Pectral Analysis with **F**requency **D**ependent **R**esolution)
 - § metóda ETFE (**E**mpirical **T**ransfer **F**unction **E**stimate)
- identifikácia modelov s impulznou charakteristikou
 - § korelačné metódy
- identifikácia modelov s prenosom
 - § metóda prenosu PEM
- identifikácia vstupno-výstupných polynomiálnych modelov
 - § metóda ARX
 - § metóda ARMAX
 - § metóda Box-Jenkins
 - § metóda Output-Error
- identifikácia modelov v stavovom priestore (modely s jedným vstupom, modely s viacnásobným vstupom, modely s viacnásobným výstupom)
 - § metóda N4SID
 - § metódy podpriestoru
 - § metóda PEM

- **Nelineárne modely identifikácie**

- nelineárne ARX modely (modely vstupno-výstupných dát v časovej oblasti alebo dát časových radov, modely s jedným alebo s viacnásobným vstupom)

- § nelineárna metóda ARX

- Hammerstein-Wiener-ove modely (modely vstupno-výstupných dát v časovej oblasti)

- § Hammerstein-Wiener-ova metóda

- **Modely identifikácie časových radov**

- identifikácia modelov s frekvenčnou charakteristikou

- § metóda SPA (**S**Pectral **A**nalysis – Blackman-Tukey)

- § metóda SPAFDR (**S**Pectral **A**nalysis with **F**requency **D**ependent **R**esolution)

- § metóda ETFE (**E**mpirical **T**ransfer **F**unction **E**stimate)

Samotnému procesu identifikácie v technologických procesoch predchádzajú merania údajov, zber a reprezentácia nameraných hodnôt, výber relevantných údajov a ich predspracovanie. Keďže prostredie MATLABu poskytuje nástroje pre sprístupnenie databázového rozhrania k údajom reprezentovaným prostredníctvom multidimenzionálnej dátovej štruktúry, je možné v procese identifikácie využiť široké možnosti spracovania údajov a dátovej analýzy, ktoré poskytuje dátový sklad. V prvom rade sa jedná o využitie robustných nástrojov určených pre proces ETL, ktoré zabezpečia automatizovaný spôsob ako sprístupniť technologické údaje z nehomogénnych zdrojov, transformáciu údajov do relevantnej podoby a načítanie údajov do adekvátnych dátových štruktúr. Ďalšou zásadnou možnosťou je využitie techník dolovania dát, ktoré by boli schopné odhaliť nové, klasickými metódami identifikácie neodhalené závislosti medzi meranými veličinami. Vhodné metódy pre odhaľovanie nových vzťahov v objemných dátach sú napr. lineárna a nelineárna regresná analýza, ktoré majú veľký význam pre predikciu hodnôt náhodných veličín na základe znalostí iných veličín, metódy predikcie v časových radoch, kde techniky dolovania dát okrem Box-Jenkinsovej metodológie poskytujú analýzy založené na báze neurónových sietí. Pre analýzu vplyvov jednotlivých veličín techniky dolovania dát disponujú metódami diskriminačnej analýzy, logistickej regresnej analýzy a tiež metódami, ktoré sú založené na princípe rozhodovacích stromov a neurónových sietí. (3, 4)

5. Zhrnutie

V príspevku sme sa pokúsili poukázať na využitie technológie dátových skladov a dolovania, ktoré poskytujú široké možnosti zberu a reprezentácie údajov a poskytujú techniky vhodné na získavanie znalostí o vzájomných dátových interakciách, ktoré je možné využiť v identifikačnom procese. Pokúsili sme sa navrhnúť optimálnu multidimenzionálnu štruktúru pre reprezentáciu hodnôt nameraných veličín pre mnohorozmerný dynamický systém. Taktiež sme predstavili identifikačné možnosti nástroja pre vedecko-výskumné výpočty a naznačili možnosti spolupráce MATLABu s databázovým rozhraním pre prístup k multidimenzionálnym štruktúram dátového skladu.

6. Zoznam bibliografických odkazov

- (1) Bruckner, R. M., Jeng, J. J., Schiefer, J.: Real-time Workflow Audit Data Integration into Data Warehouse System. ECIS, Naples, 2003.
On-line [<http://en.scientificcommons.org/50066073>], [cit.:26.10.2010]
- (2) Adelman, S., Moss, L. T.: Data warehouse project management. Addison-Wesley Longman Publishing Co., Inc. Boston, 2000. ISBN 13: 978-0-201-61635-4
- (3) Hudzovič, P.: Identifikácia a modelovanie. Bratislava, ES SVŠT, 1986.
- (4) Söderström, T., Stoica, P.: System identification. Prentice Hall Int., 1989. ISBN: 0-138-81236-5

7. Adresy autorov:

Michaela Horalová Kalinová, Mgr.
Katedra aplikovanej informatiky FPV UCM
Nám. J. Herdu 2
917 01 Trnava
michaela.horalova@ucm.sk

Stanislav Horal, Mgr.
Katedra aplikovanej informatiky FPV UCM
Nám. J. Herdu 2
917 01 Trnava
stanislav.horal@ucm.sk