

Získavanie znalostí z databáz a ich využívanie vo vizualizácii informačných a riadiacich systémov (1)

Marek Štofko, Eva Ocelíková

Cielom príspevku je poukázať na možnosť získavania znalostí z časovo usporiadaných množín údajov (databáz) ako prostriedku na podporu rozhodovania v oblasti automatizácie a v informačných a riadiacich systémoch (IRS). Uvedené sú základné informácie o probléme rozhodovania a jeho podpore. Prezentujeme definíciu získavania znalostí z databáz a dolovania údajov. Pretože oblasť automatizácie sa vyznačuje zberom časovo následných údajov, jedným z prístupov pri získavaní znalostí z takýchto databáz je hľadanie frekventovaných epizód v sekvenciách udalostí. Aplikácia tohto prístupu na údaje z reálnych procesov so sebou prináša rôzne problémy, napr. definovanie množiny možných udalostí. Tomuto problému sa venuje posledná časť príspevku.

Úvod

Neustále zvyšovanie rýchlosti počítačov, zväčšovanie kapacity diskových médií a pamätí, možnosť rýchleho prístupu k archivovaným údajom, ako aj možnosť vzdialeného prístupu k údajom cez internet, to všetko vedie k neustále sa zvyšujúcemu nároku na schopnosti ľudského vnímania. Dôsledkom toho je snaha o vytvorenie podporných softvérových prostriedkov na zobrazovanie rôznych riadiacich i riadených systémov a toku údajov v reálnych procesoch v reálnom čase. Ich cieľom je sprístupniť človeku informácie vo vhodnej forme na podporu rozhodovania v automatizovanom procese riadenia. Napriek tomu obrovské množstvo údajov snímaných a zaznamenávaných počas prevádzky rôznych systémov nie vždy umožňuje postrehnúť ich vzájomné vzťahy a možné informácie ukryté v nich. Tento problém sa snažia riešiť nové moderné prístupy. Jedným z nich je aj získavanie znalostí z databáz (Knowledge Discovering in Database – KDD).

1. Podpora rozhodovania s využitím KDD v IRS

Štruktúru informačného riadiaceho systému (IRS) možno znázorniť pomocou pyramídového modelu. Počet úrovní pyramídového modelu závisí od zložitosti systému. Tieto modely sú bližšie opísané v [8], [9]. Na obr. 1 je znázornený štvorúrovňový model IRS.

Základné delenie IRS vzhľadom na typ riadenia:

- automatické – úloha riadenia sa kladie na automat,
- automatizované – v tomto prípade zasahuje do riadenia aj človek – operátor.

Úrovne modelu IRS:

- technologická – najnižšia úroveň samotného technologického procesu,
- vizualizačná – stredná úroveň, týka sa operátorov, dispečerov, výskumných pracovníkov, kde sa prejavuje interakcia s nižšou úrovňou v pohľade na kvalitu technologického procesu,
- informačná – na túto najvyššiu úroveň postupujú údaje z nižších úrovní, spracované vo vhodnej forme na kontrolu ekonomického úžitku prebiehajúceho procesu vzhľadom na produkciu a spotrebu.

S narastajúcou zložitou technologických a informačných procesov a expanziou objemov údajov narastá aj zložitnosť a komplexnosť problémov, vznikajúcich v procese rozhodovania. V automatizovaných systémoch riadenia vykonáva rozhodovanie operátor,

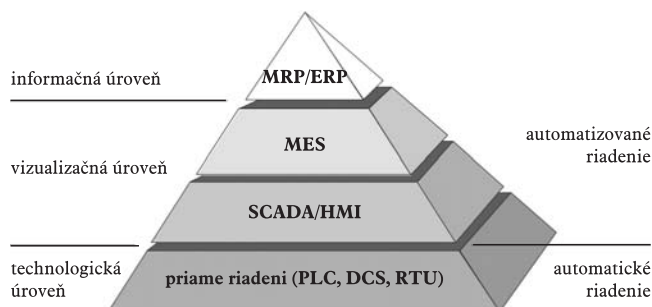
dispečer, tzv. účastník rozhodovacej situácie alebo automatizovaný riadiaci člen, tzv. formátor [6].

Zo systémového hľadiska možno pojem rozhodnutie definovať ako výsledok spoločných znakov rôznych analýz [5]: „Rozhodnutie je cielavedomou voľbou medzi verziami činností v danom prostredí, kde sa rôzne verzie činností v procese rozhodovania v čase pred rozhodnutím prejavujú ako možnosti činností.“

S výrazmi ako optimálne rozhodnutie, optimálny variant, najvhodnejšie riešenie a pod. sa stretávame aj v bežnom živote. Tieto pojmy však nie sú jednoducho definovateľné, pretože závisia najmä od situácie, v ktorej sa rozhoduje. Výber optimálneho rozhodnutia je relatívne jednoduchý, ak účastník prijímajúci rozhodnutie [3]:

- pozná všetky možné verzie činností,
- s určitou vierou, aké výsledky prinášajú jednotlivé verzie činností,
- je schopný presne určiť preferenčné poradie výsledkov.

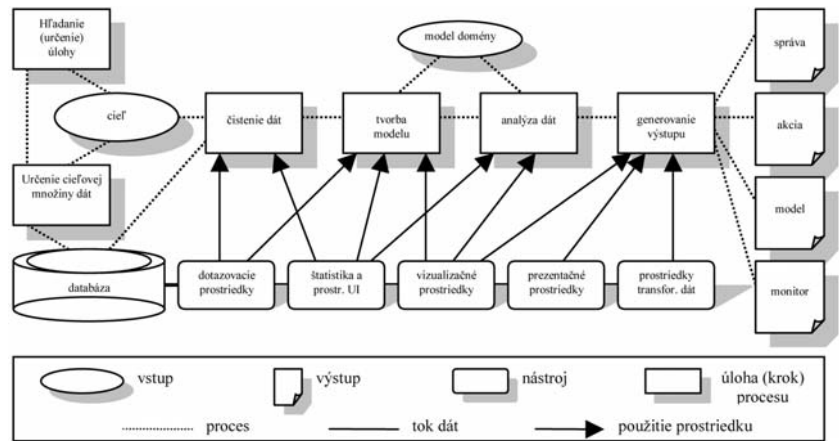
Platnosť týchto troch podmienok súčasne je málokedy splnená. Väčšinou ide o veľké množstvo informácií, ktoré by musel človek zohľadniť, aby mohol vôbec určiť množinu možných riešení a nie ešte „najvhodnejšie“ riešenie. Na to sa vytvorili metódy na podporu rozhodovania, ktoré umožňujú určiť množinu možných riešení, prípadne zhodnotiť vhodnosť možných riešení pre danú situáciu. Samotný výber a uskutočnenie rozhodnutia je v mnohých prípadoch ponechané na účastníka rozhodovacej situácie (operátora, dispečera). Teda cieľom je uplatniť podporu rozhodovania s využitím informácií získaných z údajov v tých častiach IRS, kde človek zasahuje do procesu. Dá sa povedať, že okrem technologickej úrovne (obr. 1) na všetkých vyšších úrovniach má hlavné slovo človek (operátor, dispečer, inžinier, manažér).



Obr.1 Pyramídový model Informačného riadiaceho systému (IRS)

Možnosti využitia informácií získaných z údajov na jednotlivých úrovniach IRS:

- Supervisor Control and Data Acquisition/ Human-Machine Interface (SCADA/HMI) – operátor, dispečer:
 - na základe dát z technologickej úrovne, je pomocou dolovania dát možné opisovať aktuálny stav a správanie výrobného systému, predikovať havarijné stavy, anomálie vo výrobe a riadiť kvalitu.
- Manufacturing Execution System (MES) – výskum, inžinieri:
 - interakcie medzi človekom a strojom (HMI) predstavujú údaje, z ktorých možno získať informácie napomáhajúce pri riešení nasledujúcich problémov:
 - ako navrhnuť HMI systémy, aby bola komunikácia človek – stroj čo najefektívnejšia,
 - akým spôsobom inštruovať a podporovať rozhodovanie v kritických situáciách,
 - ako pristupovať k školeniu a výučbe operátorov, dispečerov a odborných technikov.
- Manufacturing Resource Planning (MRP) – manažment, ekonómia:
 - získané informácie sú „podobné“ ako informácie z bánk a supermarketov, kde ide o vyhodnocovanie a analýzu dopytu, predaja a nákladov vzhľadom na spotrebu a produkciu.



Obr.2 Proces získavania znalostí z databázy

- Tvorba modelu a analýza údajov sú komplementárne a často dochádza ku kmitaniu medzi týmito krokmi.
- Generovanie výstupu (Output Generation) – výstup môže nadobúdať rôzne formy (obr. 2). Najvýhodnejšie by bolo získať na výstupe návrh akcie. Požadovaný druh výstupu môže byť základom špecifikácie vytváratej aplikácie KDD.

2. Proces získavania znalostí z databáz a dolovanie údajov

Skutočnosť, že údaje môžu obsahovať cenné informácie na podporu rozhodovania, ovplyvnil vývoj nových prístupov a technológií na podporu rozhodovania. Jedným z prístupov je aj získavanie znalostí z databáz (Knowledge Discovering in Database – KDD) a dolovanie údajov (Data Mining – DM).

Získavanie znalostí z databáz (KDD) a dolovanie údajov (DM) je definované podľa Fayyada [2] takto: „KDD je netriviálny proces rozpoznávania platných, nových (neobvyklých), potenciálne užitočných a jednoznačne zrozumiteľných vzorov z údajov.“ Ďalej sa uvádza, že objavovanie znalostí v databáze je interaktívny a iteratívny proces s niekoľkými krokmi a dolovanie údajov je jeho časťou (jedným krokom). Podrobnejšie definície pojmov sa nachádzajú aj v [7]. Schéma KDD procesu podľa [1] je na obr. 2.

Opis procesu KDD, znázorneného na obr. 2:

- Hľadanie úlohy (Task Discovery) – potenciálny konečný používateľ formuluje cieľ, ktorý sa zdá jasný. Až po hlbšom skúmaní formulovaného cieľa sa presne a jasne sformuluje úloha.
- Určenie cieľovej množiny údajov (Data Discovery) – je doplnujúci krok k predošlému kroku hľadania úlohy a obsahuje zhodnotenie, či je kvalita údajov dostatočná na riešenie formulovanej úlohy.
- Model domény (Domain Model) – má dôležitú úlohu pri získavaní znalostí z databázy, väčšinou však zostáva iba v mysli analytika (experta).
- Čistenie údajov (Data Cleaning) – je náročná úloha ďalšieho kroku procesu KDD. Môže sa stať, že pri čistení sa odstránia údaje, ktoré sa javia ako zašumené a pritom to môže byť anomálie, ktorú by bolo výhodné preskúmať.
- Tvorba modelu (Model Development) – je kľúčovou časťou procesu KDD, predchádza analýze údajov. Analytik je po kontakte s údajmi schopný sformulovať hypotézy (využíva svoje znalosti z danej oblasti).
- Analýza údajov (Data Analysis) – snaha pochopiť, prečo sa určité skupiny entít správajú tak, ako sa správajú; ide o hľadanie zákonitostí určitého správania.

Literatúra

- [1] BRACHMAN, R. J., ANAND, T.: The Process of Knowledge Discovery in Databases. Eds. AAAI/MIT Press, Cambridge, Massachusetts, 1996.
- [2] FAYYAD, U. M., PIATETSKY - SHAPIRO, G., SMYTH, P.: The KDD Process for Extracting Useful Knowledge from Volumes of Data. Communication of the ACM, Nov. 1996, Vol. 39, No. 11, pp. 27 – 34.
- [3] MADARÁSZ, L.: Inteligentné technológie a ich aplikácie v zložitých systémoch. University Press elfa, Košice 2004, s. 345.
- [4] MARK, B.: Data mining – Here we go again? IEEE EXPERT, Oct. 1996, Vol. 11, No. 5.
- [5] OCELÍKOVÁ, E., ŠTOFKO, M.: Nové prístupy v oblasti podpory rozhodovania v informačných a riadiacich systémoch. In: Proc. of International Conference Cybernetics and Informatics. April 5. – 6. 2001, Piešťany, pp. 57 – 58.
- [6] ŠTOFKO, M., OCELÍKOVÁ, E., ZOLOTOVÁ, I.: Získavanie znalostí z databáz časovo následne usporiadaných dát na podporu rozhodovania v IRS. In: Proc. of International Conference „Znalosti '01“, Jun 19 – 21, 2001, Prague, Czech Republic, pp. 124 – 134.
- [7] ZOLOTOVÁ, I., ŠTOFKO, M.: Simulácia a procesná vizualizácia pre podporu rozhodovania. In: Proceedings of 33rd Spring International Conference, Modelling and Simulation of Systems, MOSIS '99, Volume 1, Rožnov pod Radhoštěm. Apríl 1999, pp. 55 – 62.

Pokračovanie v budúcom čísle.

prof. RNDr. Eva Ocelíková, CSc.
Ing. Marek Štofko

Technická univerzita v Košiciach
Fakulta elektrotechniky a informatiky
Katedra kybernetiky a umelej inteligencie
Letná 9, 041 20 Košice
e-mail: Eva.Ocelikova@tuke.sk
marek_stofko@hotmail.com

29

