

Inteligentní řízení a rozhodování v distribučních sítích

Jan Kout, Tomáš Hejny, Jiří Kléma

Distribuční sítě jsou komplexní a dynamické systémy. Pro jejich efektivní řízení je třeba důkladně porozumět základním mechanismům jejich chování. Správné zodpovězení otázek typu *kolik? kdy? kam?* přináší distributorům mnoho finančních výhod. Obecně existuje několik způsobů, jak stále nenahraditelné lidské porozumění doplnit porozuměním strojovým. Článek prezentuje řešení založené na podpoře rozhodování, realizované jako součást informačního systému distribuční společnosti. Bloky pro podporu rozhodování a řízení jsou tvořeny prediktivními a deskriptivními modely, automaticky generovanými moderními metodami dolování dat. Modely jsou vytvářeny v rámci predikčního systému, jenž je výsledkem spolupráce mezi výzkumným oddělením firmy CertiCon, a. s., a Gerstnerovou laboratoří pro inteligentní rozhodování a řízení ČVUT Praha.

1. Motivace, problémy

Distribuční sítě komodit jako elektřina, voda, plyn nebo teplo jsou prostředím vyžadujícím rychlé a efektivní řízení a rozhodování. Lidský operátor na základě své zkušenosti dobře chápe základní principy určující chování sítě. Současně ale jde o prostředí dynamická a hierarchická, generující velké objemy časových dat. Jde tedy o prostředí, kde může docházet k chybným nebo pozdním reakcím na konkrétní, potenciálně ne zcela obvyklý vývoj událostí. Adaptivní systémy pro podporu rozhodování založené na automaticky vytvářených datových modelech mohou významně napomoci kvalitě rozhodování a řízení. Obvykle jde o minimalizaci distribučních ztrát včasnou reakcí na změnu vstupních, vnitřních či výstupních parametrů sítě. Důsledkem včasných a přesných řídicích zásahů je také optimalizace nákladů odvozených ze smluvních vztahů, a tím i snížení koncové ceny předmětu spotřeby, zvýšení provozního zisku či komfortu koncových odběratelů.

Moderní metody strojového učení, tzv. dolování dat (získávání informací ze souborů dat – viz dále) a matematické statistiky nabízejí množství možností, jak zmíněné modely vytvářet. Z obecného pohledu jde o induktivní postupy založené na zobecnění konkrétních, systematicky a dlouhodobě sledovaných údajů. Základními údaji jsou data zjišťovaná na různých místech distribuční sítě – typicky jde o odběr nebo průtok v daném kontrolním bodě. Doplnkovými údaji mohou být data zjišťovaná mimo distribuční síť a mající úzký vztah k objemu spotřeby sledované komodity – nejčastějšími doplnkovými veličinami jsou údaje o počasí,

dále lze uvažovat např. cenu distribuovaného předmětu spotřeby či vstupních surovin.

Sledované veličiny jsou vzorkovány po určitých, nejčastěji pevně stanovených časových intervalech. Modelování v distribučních sítích lze převádět na analýzu, resp. „dolování“ z časových řad více proměnných. Teoretickým otázkám zpracování časových dat budou věnovány další odstavce. Na tomto místě uvedeme typizované příklady problémů z praxe, které můžeme řešit. Zaměříme se pro tuto chvíli na hodinový odběr vody v jedné čtvrti. V první řadě se můžeme snažit analyzovat danou časovou řadu přímo v časové oblasti. Hledáme v jejím rámci charakteristické podsekvence (najdi den, kdy se odběr choval stejně jako dnes, popř. najdi dny nebo hodiny s netypickým průběhem spotřeby), porovnáváme průběhy spotřeby s jinými řadami (najdi lokality, ve kterých je charakter spotřeby podobný jako u cílové lokality) nebo se snažíme ověřit stacionaritu sledovaného systému (nedochází k posunu konceptu, tj. časové změně schématu chování sledovaného systému?). Z důvodů možných časových posuvů, zrychlování a zpomalování změn v čase a také zašumění se často nejedná o triviální otázky.

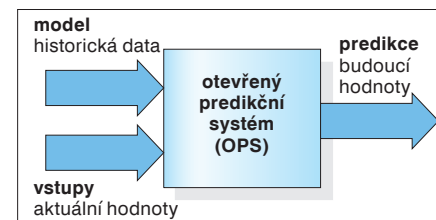
Nejobvyklejším problémem je však nalezení parametrů (atributů), které danou časovou řadu co nejlépe charakterizují. Vraťme se k našemu příkladu a pokusme se nalézt takové parametry, které nám každé ráno umožní co nejpřesněji odhadnout hodnotu spotřeby večer stejného dne. Můžeme spoléhat na všechna data, která máme k dispozici v daný den ráno, tj. jediným omezením je, že nesmíme využívat tzv. anachronistické údaje. Předpokládejme, že máme k dispozici údaje o spotřebě vody z dalších sídlišť, údaje o počasí či přehled zajímavých kulturních a jiných masových akcí v dané lokalitě. Výsledkem analýzy je seznam veličin, které vykazují asociaci s cílovou veličinou (např. úhrn srážek za předešlý den, roční období reprezentující jistou setrvačnost spotřeby, zajímavost večerního kulturního programu a také spotřeba ve stejnou hodinu včerejšího večera). Časovou řadu potom reprezentujeme vektorem atributů a nad takto vzniklým prostorem definujeme podobnostní metriku, prostorové indexy apod.

Získané informace je možné zobrazit v podobě popisného, tj. deskriptivního modelu. Tento model operátorovi prezentuje netriviální znalosti o dané síti. Jednoduchým příkladem je srovnání relativní důležitosti jednotlivých popisných parametrů v různých částech sítě (čtvrť A reaguje hodně citlivě na včerejší srážky, zatímco čtvrť B je citlivější na to, jaký

je den v týdnu). Z nich lze odvozovat řídicí zásahy v dané síti. Druhou možností je vytváření prediktivních modelů, tj. modelů snažících se podat kvantitativní předpověď cílové veličiny. Dané předpovědi mohou operátorovi opět sloužit k optimalizaci řídicích akcí. Další text bude zaměřen právě na prediktivní modely.

2. Dolování dat, strojové učení a predikční problémy

Takzvané dolování dat je chápáno jako hledání cenných informací ve velkém množství zdrojů a dat [8]. Tyto informace mohou být různě prezentovány a mohou mít velmi rozsáhlé možnosti využití v praxi. Jednou z častých oblastí aplikací je nalezení modelů, které jsou vhodné pro predikci (obr. 1).

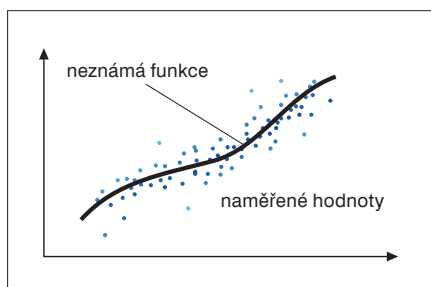


Obr. 1. Základní princip modelování

Predikce je určování současných či budoucích stavů a hodnot, které nejsou vzhledem ke složitosti a komplexnosti pozorovaného objektu běžnými způsoby odhalitelné či měřitelné. V uvažovaných datech neexistuje žádná zřejmá a přesně definovatelná závislost mezi hodnotami v minulosti a budoucnosti. Právě metody umělé inteligence umožňují řešit i takovéto úlohy. Dolování dat je kooperativní činnost lidí a počítačů, vybavených vhodným softwarem. Existuje mnoho algoritmů strojového učení a statistických metod, které lze využít k dosažení stanovených cílů. Strojové učení může být také chápáno jako počítačový prostředek, který se učí ze zkušenosti (minulosti) vzhledem k definici a cílům úlohy.

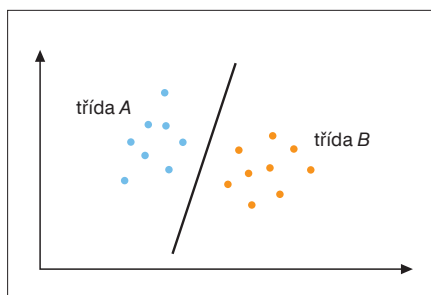
V prediktivních úlohách uživatele zajímá hodnota konkrétní, tzv. cílové veličiny, typicky relačního atributu, v závislosti na jiných, známých veličinách (atributech). Prediktivní modely mají za cíl předpovídat neznámou nebo dosud neměřenou hodnotu na základě jiných známých hodnot charakterizujících objekt predikce. Pro každou konkrétní úlohu je množina uvažovaných veličin jednoznačně rozdělena na veličinu závislou (výstup) a veličiny nezávislé (vstupy, atributy). Existují dva základní typy predikčních úloh: *regrese* a *klasifikace*.

Regrese je problém odhadování neznámých funkcí či hodnot (např. zítější spotřeba elektřiny bude 250 MW-h; obr. 2). Regresní model je vytvářen na základě číselných vstupů a jim odpovídajících výstupů. Výstupem regresní predikce je číselná hodnota reprezentující např. budoucí množství a trendy.



Obr. 2. Regresní úloha

Klasifikace je problém hledání skrytých stavů (obr. 3). Typicky je třeba odhalit současný či budoucí stav určitého systému (bezpečný či nebezpečný stav), který nelze přímo změřit. K dispozici je však množství měření a dat, která s daným stavem souvisejí. Podle výsledků těchto měření a příslušného stavu je vytvořen model. Spolu s vytvořeným modelem je odhadnuta jeho přesnost predikce. Ta určuje možnosti dalšího použití.



Obr. 3. Klasifikační úloha

Obecně platí, že rozhodnutí jsou činěna na základě informací a zkušeností. Čím více spolehlivých informací je k dispozici, tím relevantnější rozhodnutí může být učiněno. Kvalita dosaženého výsledku je dána volbou vhodného modelu. Současně však musí platit to, že model je v průběhu učení konfrontován s dostatečným množstvím vhodně předzpracovaných údajů, které jsou platné (tj. nejsou zastaralé), spolehlivé (minimalizuje se výskyt chybějících a nesprávných hodnot) a pokrývají celou nebo alespoň významnou škálu podmínek, za kterých bude model v budoucnu používán (obecně platí, že techniky interpolace jsou spolehlivější než extrapolací techniky).

3. Příklady reálných úloh

3.1 Lokální distributor plynu – predikce denní spotřeby plynu

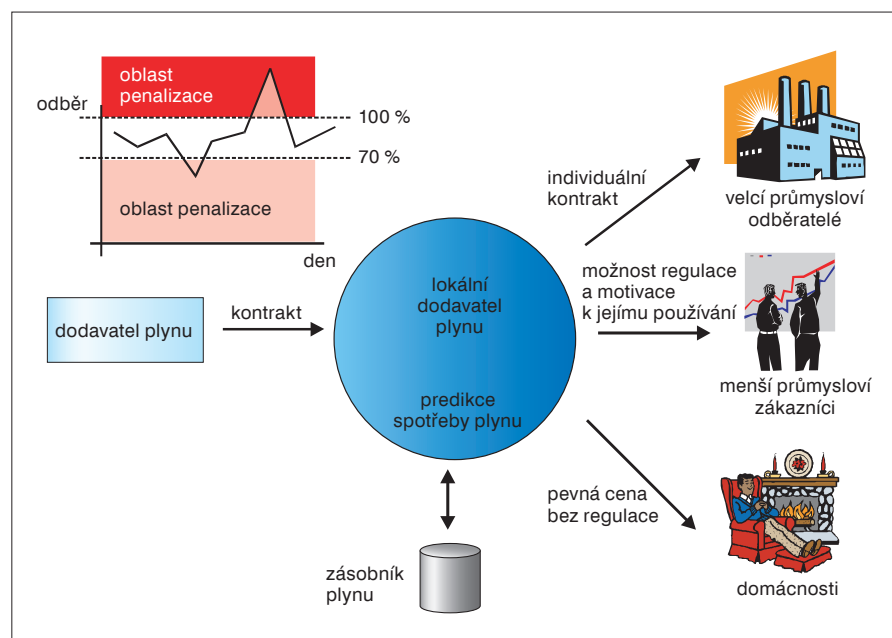
Lokální distributoři plynu zpravidla odebírají plyn od národních dodavatelů a dále jej distribuují v rámci regionu, který pokrývají.

Jejich kontrakt s národními dodavateli je obvykle limitován a určen smluvními denními limity odběru. Lokální distributoři musí každodenně udržovat své odběry v těchto předem dohodnutých mezích. Kdykoliv je limit překročen, lokální distributor musí platit vysokou pokutu. Je zřejmé, že denní predikce množství odebraného plynu může lokálním distributorům značně napomoci regulovat svůj odběr v rámci smluvených mezí. Zpravidla mají dvě možnosti, jak se nežádoucí odběrové špičky a z ní plynoucí penalizaci vyhnout. Za prvé mohou některým svým odběratelům po dohodě doporučit přepnutí na alternativní zdroj energie (např. elektřinu či olej). Druhá možnost, jak předjet penalizaci, je naplnit vybudované rezervní

lze překlenout očekávané výkyvy v odběru. Dlouhodobé předpovědi (týdny) lze využít k optimalizaci zásob vstupního média.

4. Základní požadavky na predikční systém

Úlohy modelování distribučních sítí je možné řešit dvěma principiálně odlišnými způsoby: predikcí na základě fyzikálního modelování systému nebo predikcí na základě datového modelování (strojové učení). Fyzikální modely se obecně uplatňují u jednodušších a identifikovatelných problémů. Základní požadavky na datově orientovaný predikční



Obr. 4. Predikce spotřeby plynu

nádrže plynu během období menších spotřeb. Zásoby z nádrží se potom použijí během dnů se zvýšeným odběrem plynu (obr. 4).

Přesnost predikce je závislá na typu sítě, odběratelích i kvalitě údajů nasbíraných v minulosti. Přesnost predikce v zimních měsících bývá zpravidla od 96 do 99 %. Predikci je možné během dne kontinuálně zlepšovat a zpřesňovat.

3.2 Distributor tepla – predikce spotřeby tepla

Výroba a dálková distribuce tepla pomocí ohřevu a distribuce tepelného média (vody) jsou závislé na vstupním médiu, které nese teplo. Množství odebraného tepla je značně ovlivňováno venkovními teplotami. Vztah není triviální s ohledem na relativně dlouhé a různé časové konstanty, které se při odběru tepla uplatňují. Rozvod tepla rovněž závisí na topologii sítě, jeho ztrátách a typu odběratelů. Krátkodobé předpovědi spotřeby tepla (v řádech dnů) se uplatňují při optimalizaci ohřevu (teploty) distribuovaného média. Jeho korekcemi

systém pro problémy časových řad více proměnných lze shrnout do zhruba tří oblastí.

Za prvé, systém musí mít schopnost flexibilně předzpracovávat a upravovat vstupní data; to je vzhledem k jejich objemu a chybějícím či chybným údajům klíčová vlastnost. Uvádí se, že až 80 % úsilí bývá věnováno právě předzpracování a manipulaci se vstupními daty. Datové zpracování je podpořeno analýzou dat a hledáním vhodných atributů ze vstupních dat. Výběr vhodných atributů slouží k vytváření modelů.

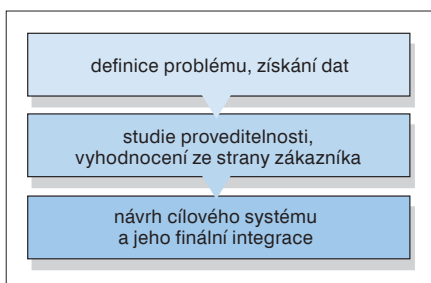
Za druhé, je třeba rychle vytvářet predikční modely. K tomu patří i možnost testování modelů a simulace reálného provozu. Zvláště simulace reálného provozu s daty, která nebyla použita při analýze dat a vytváření modelů, reálně zhodnocuje přesnost a použitelnost predikčního systému.

Za třetí, důležitým faktorem je rychlá integrace a zapojení predikčního systému do informačních systémů či jeho zavedení jako samostatného softwarového produktu. Tato fáze představuje přenesení navrženého řešení do cílového prostředí.

5. Návrh a vývoj predikčního modulu

Vývoj predikčního modulu pomocí metod strojového učení lze obecně shrnout do tří základních fází. Ukazuje je *obr. 5*.

První fáze, definice problému, analyzuje problém, nastavuje kritéria úspěšnosti řešení a vybírá relevantní data. Vybraná data a jejich kvalita určují úspěšnost řešení problému. Základním kritériem je jejich relevantnost a množství. Čím více relevantních dat je

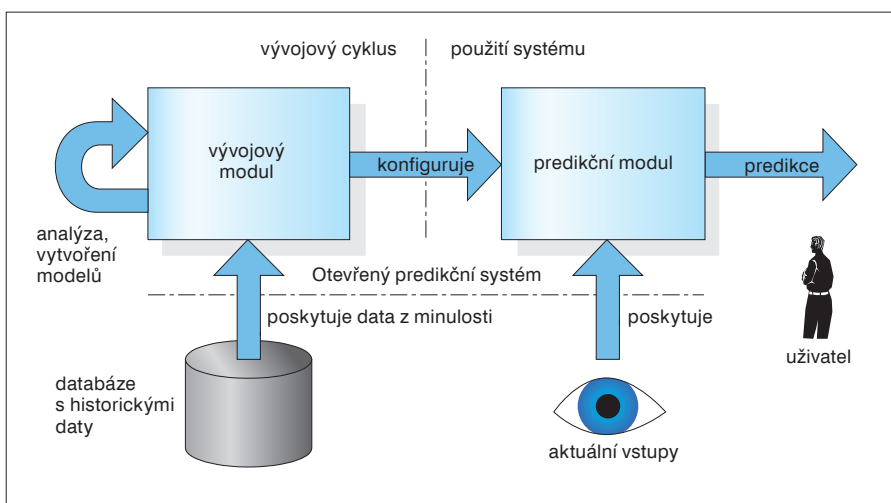


Obr. 5. Fáze návrhu a vývoje predikčního systému

k dispozici, tím lépe je možné vytvářet modely a simulovat a testovat reálný provoz. Důležitým vstupem je také navržené kritérium úspěšnosti navrženého řešení. V této fázi se předpokládá úzká spolupráce mezi dodavatelem dat a tvůrcem predikčního systému.

Studie proveditelnosti slouží k analýze dat a návrhu řešení systému s předpokládanou přesností. Výstupem studie proveditelnosti je prototyp predikčního modulu, definice vstupů a výstupů a předpokládaná přesnost predikce.

Cílový systém a jeho implementace (*obr. 6*) jsou opět navrhovány v úzké spolupráci s expertem. Jsou dvě základní možnosti řešení cílového systému: samostatný softwarový produkt či integrace do již existujícího informačního systému. Nedílnou součástí implementace je ověření v reálném provozu. Takto navržený predikční systém je možné v libovolných intervalech aktualizovat s využitím nově nasbíraných dat. Některé algoritmy vedle toho mají schopnost se autonomně aktualizovat.



Obr. 6. Vývoj predikčního systému

6. Otevřený predikční systém

Otevřený predikční systém (OPS) představuje možnou variantu řešení predikčních úloh. Jedná se o softwarový nástroj založený na prediktivní metodologii. Zahrnuje zpracování dat, toky a prediktivní algoritmy [2]. Prediktivní algoritmy definují modely, které jsou založeny na datech shromážděných v minulosti. Modely odhalují současné či budoucí množství, toky, stavy, chyby či jiné požadované hodnoty. OPS dále integruje prostředky analýzy a vizualizace dat a tvoří tak jednotné prostředí pro vytvoření studie proveditelnosti, návrhu systému či kompletního řešení. Hlavní zaměření systému je na:

- předvídaní problémů, které mohou vést k nebezpečným a nežádoucím situacím,
- podporu strategického plánování a alokace zdrojů,
- predikci budoucích hodnot a trendů.

Vývojový nástroj OPS nabízí několik užitečných vlastností, např.:

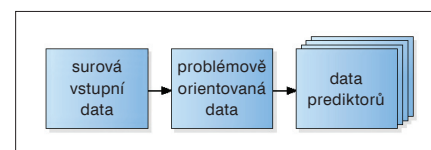
- flexibilní předzpracování dat s využitím předdefinovaného datového toku, které zahrnuje metadata, podporu datových transformací,
- podporu zpracování časových řad – taková data se často vyskytují v distribučních společnostech,
- kombinaci statistických metod s algoritmy strojového učení – podporu různorodosti řešení, s možností kombinace více přístupů vedoucí k větší stabilitě řešení,
- rychlý vývojový cyklus od dat k návrhu řešení predikčního systému.

6.1 Datové toky

Flexibilní datové toky jsou jednou z hlavních výhod navržené prediktivní metodologie (*obr. 7*, [2]). Předzpracování dat stanovuje tři základní datové struktury, které odpovídají jednotlivým krokům řešení úlohy: vstupní data, data problému a data prediktorů.

Vstupní data odpovídají datovým vstupům celého systému. Jsou většinou shodná či velmi blízká naměřeným datům odrážejícím chování

pozorovaného systému. Jejich organizace odpovídá nejkratším vzorkovaným časovým úsekům. Vstupní data jsou transformována do datové struktury, která odpovídá danému problému. Hlavním smyslem transformací je nalezení relevantních atributů pro řešení úlohy. Problémově orientovaná data (*Meta Record*) obsahují veškeré vstupy a požadované výstupy, výstupy jednotlivých prediktorů a konečné výstupy. Jednotlivé predikční algoritmy vyžadují specifický způsob předzpracování dat (např. škálování), a proto jsou data problému dále transformována do datových struktur jednotlivých prediktorů.



Obr. 7. Datové toky systému

Transformační funkce zajišťují přechody mezi jednotlivými datovými strukturami. Rozlišují se dva hlavní typy transformačních funkcí: časové a datové. Časové transformační funkce určují časovou sekvenci a zároveň umožňují časovou konverzi a kompresi (např. přechod mezi vstupními daty vzorkovanými po hodinách a úlohou, která vyžaduje týdenní predikci). Datové transformace zajišťují transformaci a předzpracování dat či hledání vhodných atributů.

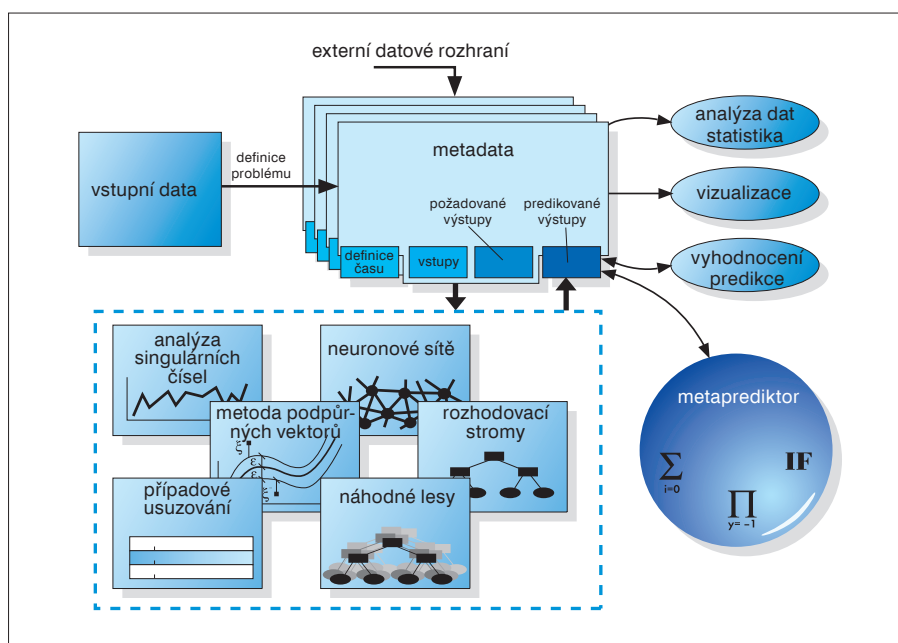
Důležitým faktorem datového zpracování je filtrace dat. Filtrace je potřebná k odstranění nežádoucích vstupů, ale zvláště se uplatňuje při vytváření lokálních modelů, které zlepšují kvalitu predikce.

6.2 Prediktivní algoritmy a jejich kombinace

Mezi hlavní vlastnosti systému patří dostupnost řady algoritmů, které se využívají k vytváření modelů (*obr. 8*). Různorodost řešení podporuje stabilitu celého predikčního systému. Mezi vhodné algoritmy patří:

- analýza singulárních čísel,
- neuronové sítě,
- metoda podpůrných vektorů,
- rozhodovací stromy,
- náhodné lesy,
- případové posuzování konfigurované pomocí genetických algoritmů,
- kombinace prediktorů.

Statisticky lze přínos kombinace více samostatných metod zdůvodnit tím, že konečné chyby jednotlivých metod lze rozdělit na chyby způsobené daty a chyby způsobené samotnou metodou. Do první kategorie chyb patří např. odchylka predikčního systému způsobená náhlou netypickou změnou vstupních hodnot, popř. jejich zcela chybným zadáním. Tyto chyby se ve výsledné predikci odrazí vždy, nezávisle na typu či kvalitě použité metody. Druhou kategorií jsou chyby způsobené metodou samou; cílem každého systému je tuto chybu minimalizovat. Vhodným způsobem



Obr. 8. Metodologie OPS

minimalizace se jeví právě kombinace více metod. Zde lze uplatnit techniku lokální komparativní výhody některé z metod. Vyplývá-li např. ze statistické analýzy úspěšnosti jednotlivých metod, že některá z nich dosahuje nejlepších výsledků za specifických podmínek (v daném měsíci, za extrémních teplot nebo v odpoledních hodinách), bude tato metoda použita přednostně právě za již zmíněných podmínek.

7. Závěr

Všeobecný rozvoj výpočetní, měřicí nebo telekomunikační techniky umožňuje organizacím s malými náklady zjišťovat a uchovávat velké objemy dat souvisejících s jejich provozem. Tato data jsou potenciálním zdrojem cenných informací. Datové sklady a s nimi spojené analytické dotazy zodpovídají jednodušší a jasně formalizovatelné otázky operativního řízení. Informatika dále poskytuje široké spektrum metod, jak dolovat strukturálně složitější asociace a odvozovat

z nich modely nejširšího použití. Distribuční společnosti jsou charakteristické typem dat, která generují, a následně způsobem jejich využití. Použití problémově orientovaných systémů dolování dat reflektujících tato specifika může výrazně usnadnit a urychlit proces získávání a využívání informací skrytých v provozních datech.

Literatura:

- [1] KLÉMA, J. – KOUT, J.: *Prediction of Gas Consumption*. In: *Systems Integration*, 99. University of Economics, Prague, 1999, p. 119–128. ISBN 80-7079-059-8.
- [2] KOUT, J.: *Prediction Methodology for Utility Companies*. Ph.D. thesis in English, ČVUT, FEL, Praha, 2001.
- [3] KOUT, J. – KLÉMA, J. – VEJMEJKA, M.: *Predictive System for Multivariate Time Series*. In: *Cybernetics and Systems 2004*. Vienna: Austrian Society for Cybernetics Studies, 2004, vol. 1.2, p. 723–728. ISBN 3-85206-169-5.
- [4] MULLER, K. – SMOLA, A. – RATSCH, G. a kol.: *Predicting Time Series with Support*

Vector Machines. ICANN '97, p. 999–1004. Springer, 1997.

- [5] PYLE, D.: *Data Preparation for Data Mining*. Morgan Kaufmann, California, 1999.
- [6] VAPNIK, V.: *The Nature of Statistical Learning Theory*. Springer NY, 1995.
- [7] WEIGEND, A. – GERSHENFELD, N.: *Time Series Prediction – Forecasting the Future and Understanding the Past*. Addison-Wesley, Reading, Massachusetts, 1993.
- [8] ŽELEZNÝ, F. – KLÉMA, J. – ŠTĚPÁNKOVÁ, O.: *Strojové učení v dobývání znalostí. Umělá inteligence (4)*. Praha, Academia, 2003, p. 355–406. ISBN 80-200-1044-0.

Jan Kout, Tomáš Hejný,
CertiCon, a. s., Aplikovaný výzkum
(kout@certicon.cz),
Jiří Kléma,
Gerstnerova laboratoř
pro inteligentní rozhodování a řízení,
katedra kybernetiky, FEL ČVUT
(klema@labe.felk.cvut.cz)

Ing. Jan Kout, Ph.D. (1972), absolvent doktorského studia se zaměřením na umělou inteligenci a bio-kybernetiku na katedře kybernetiky FEL ČVUT. V současné době vedoucí oddělení aplikovaného výzkumu firmy CertiCon, a. s., a vedoucí projektů pro společnost Vitatron Medical B.V., Grundfos A/S a další s hlavním zaměřením na aplikaci metod umělé inteligence v praxi.

Ing. Tomáš Hejný, (1975), absolvent inženýrského studia se zaměřením na umělou inteligenci na katedře kybernetiky FEL ČVUT. V současné době zaměstnanec firmy CertiCon, a. s., jako vývojář softwaru. Podílel se na vývoji predikčního systému a na několika projektech pro společnost Vitatron Medical.

Ing. Jiří Kléma, Ph.D. (1971), absolvent doktorského studia se zaměřením na umělou inteligenci a bio-kybernetiku na FEL ČVUT. Řešitel výzkumných i aplikačních projektů v oblasti strojového učení a dolování dat, mj. pro Rockwell Automation, Grundfos A/S, Abbott Laboratories, Health Data Research nebo IKEM Praha. V současné době pracuje v Gerstnerově laboratoři katedry kybernetiky FEL ČVUT.

► LonWorks na olympijských hrách

V některých sportovních areálech, v nichž se odehrávaly letošní letní olympijské hry v Aténách, byl pro propojení technických zařízení budov použit otevřený komunikační systém LonWorks. Spojuje zařízení z platformy Excel5000 (Honeywell) a měřiče elektrického výkonu WattNode (Continental Controls). Celý systém se programuje a spravuje pomocí systému CARE (Honeywell), jenž využívá síťový operační

systém LNS (Echelon). Účelem aplikace bylo monitorovat spotřebu energie a redukovat riziko výpadků napájení v průběhu her.

Měřiče WattNode a zařízení Excel5000 s certifikací LonMark jsou propojeny prostřednictvím sítě s volnou topologií a využívají jako přenosové médium spletaný dvoužilový kabel. Monitorují nízkonapěťové a vysokonapěťové rozvody a řídí přepínání na záložní zdroj v případě výpadku dodávky energie.

Výhodou komunikační sítě LonWorks je její otevřenost a možnost konfigurovat a řídit zařízení od různých výrobců jediným sí-

ťovým softwarovým nástrojem. Nejsou třeba žádné nákladné komunikační brány nebo jiná zařízení ani množství specializovaného softwaru.

Po skončení her plánuje řecká společnost Controline, realizátor projektu a místní distributor společnosti Honeywell, spojit sítě LonWorks jednotlivých sportovních zařízení navzájem s využitím sítě Ethernet. K tomu budou instalovány směrovače i LON 100. Spojení pomocí Ethernetu umožňuje využívat pro monitorování a řízení celé sítě jazyk XML a webové nástroje. (Bk)