

Milan Terek

MOŽNOSTI VYUŽÍVANIA HĽBKOVEJ ANALÝZY ÚDAJOV V EKONOMIKE¹

Abstract: *The paper deals with the characteristics of data mining. The possibilities of its using in economics are described in the present paper. Further the objectives, tasks and basic application areas of data mining are given. The multidisciplinary character of data mining is emphasized in its roots. Then, the process of data mining is described; this description offers general knowledge of application procedures. Finally, there is presented a brief overview of data mining methodologies.*

Key words: *data mining, knowledge discovery in databases, predictive data mining, descriptive data mining, tasks of data mining, process of data mining*

JEL: C 89

Úvod

V tomto článku sa pokúsime charakterizovať *hlbkovú analýzu údajov (Data Mining)* – jednu z mladých, mimoriadne dynamických disciplín, ktorá je zameraná na získavanie poznatkov z databáz, a poukázať na možnosti jej využívania. Napríklad v práci [7] sa na s. xi uvádza: „V časopise ZDNET (z 8. 2. 2001) zameranom na on line technológie sa hlbková analýza údajov hodnotí ako jeden z najrevolučnejších rozvojových smerov posledného desaťročia. V MIT Technology Review považujú hlbkovú analýzu údajov za jednu z desiatich vynárajúcich sa technológií, ktoré zmenia svet“. Aplikácie hlbkovej analýzy údajov môžu určite významne prispieť k budovaniu novej ekonomiky.

1 Charakter hlbkovej analýzy údajov

Moderná veda využíva na opísanie ekonomických, sociálnych, fyzikálnych, biologických alebo iných systémov tzv. *modely prvého princípu (first-principle models)*. Experimentálne údaje sa v tomto prístupe používajú na verifikáciu týchto modelov a na odhadovanie niektorých parametrov, ktoré je ťažké alebo nemožné

¹ Tento článok vznikol s príspevom grantovej agentúry VEGA v rámci projektu číslo 1/0437/08: *Kvantitatívne metódy v stratégii šesť sigma*.

odhadovať priamo. V mnohých situáciách sú však prvé princípy neznáme alebo študovaný systém je príliš komplexný na matematickú formalizáciu ([6], s.1).

S rastúcim využívaním počítačov sa v organizáciách nazhromaždilo veľké množstvo rozličných údajov. Keď nie sú k dispozícii modely prvého princípu, možno tieto údaje využiť na získanie modelov na odhadovanie vzťahov medzi premennými (neznámych závislostí typu vstup – výstup). Ide tu o istý odklon od klasického modelovania a analýzy, založených na prvých princípoch, k tvorbe modelov a príslušných analýz priamo z údajov. Schopnosť získavať poznatky skryté v týchto údajoch má v súčasnosti stále rastúci význam.

Všeobecne – proces aplikácie metodológie založenej na počítačoch, ktorá zahŕňa nové techniky na získavanie poznatkov z údajov, sa nazýva *hlbková analýza údajov* ([6], s.2).

2 Získavanie poznatkov z databáz a hlbková analýza údajov

Začiatkom 90. rokov 20. storočia sa začalo hovoriť o získavaní poznatkov z databáz² (*KDD – knowledge discovery in databases*). Bola prijatá takáto definícia disciplíny: Získavanie poznatkov z databáz je *netriviálne získavanie implicitných, predtým neznámych a potenciálne užitočných informácií z údajov* [4]. Hlbková analýza údajov sa považovala za jeden krok procesu získavania poznatkov z databáz – krok, v ktorom sa aplikujú techniky hlbkovej analýzy údajov na odhadovanie modelu systému. Túto terminológiu ponúkol U. Fayyad na prvej medzinárodnej konferencii o získavaní poznatkov z databáz a o hlbkovej analýze údajov, ktorá sa konala v roku 1995 v Montreale, a postupne ju prijali mnohí autori, napr. [1], [2].

Postupne sa termín hlbková analýza údajov stal synonymom celého procesu získavania poznatkov z databáz, napr. v [5] a [6]. Takto ho budeme chápať aj v tomto článku.

Hlbkovú analýzu údajov možno definovať ako proces výberu, skúmania a modelovania veľkého množstva údajov, zameraný na objavovanie vzorov alebo vzťahov, ktoré predtým neboli známe, s cieľom získať jasné a užitočné výsledky pre vlastníka databázy ([5], s. 2).

Všeobecne ide o interaktívny a iteratívny proces, ktorý tvoria kroky špecifikácie problému a príslušných hypotéz, zhromaždenia údajov, ich spracovania, odhadovania modelu a interpretácie výsledkov a formulácie záverov ([6], s. 6–8). Niektorí autori zasa uvádzajú až sedem krokov ([5], s. 6–10), niektorí šesť krokov ([1], s. 37). V procese hlbkovej analýzy údajov sa kladie dôraz aj na prípravu údajov na analýzu a na interpretáciu získaných poznatkov. Pri príprave údajov sa obyčajne z údajov uložených v zložitej štruktúre vytvára tabuľka, ktorá obsahuje relevantné údaje. Štandardným modelom štruktúrovaných údajov pre hlbkovú analýzu údajov

² Niektorí autori používajú pojem *objavovanie poznatkov v databázach*.

je súbor prípadov,³ ktoré sú charakterizované pomocou hodnôt množiny znakov. Údaje pre aplikácie hĺbkovej analýzy údajov sú obyčajne usporiadané v tabuľke, v ktorej sú v riadkoch prípady a v stĺpcoch znaky. Prípady môžu byť napríklad zákazníci banky, ktorí sú charakterizovaní hodnotami množiny znakov.

Množiny prípadov sa rozdeľujú na *tréningovú* a *testovaciu* množinu prípadov. Tréningové údaje sa využívajú vo fáze vlastného učenia sa, testovacie údaje sa používajú na preverenie získaných znalostí. Niekedy sa navyše pracuje aj s *validačnými* údajmi, ktoré sa používajú na prípadné modifikácie znalostí získaných na základe tréningových údajov.

Získané poznatky sa pri interpretácii hodnotia z hľadiska finálneho používateľa.

3 Ciele hĺbkovej analýzy údajov

V hĺbkovej analýze údajov sa najčastejšie sleduje jeden z týchto cieľov: *predikcia* alebo *opis*. Predikcia spočíva vo využití premenných na predikciu neznámych hodnôt iných premenných. Opis spočíva v hľadaní *vzorov (patterns)*, ktoré opisujú údaje a ktoré možno interpretovať.

Podľa sledovaného cieľa možno potom aktivity hĺbkovej analýzy údajov zaradiť do jednej z týchto kategórií ([6], s. 2):

- prediktívnej hĺbkovej analýzy údajov, ktorá poskytuje model systému opísaného danou množinou údajov,
- opisnej hĺbkovej analýzy údajov, ktorá poskytuje nové informácie založené na danej množine údajov.

Model, ktorý poskytne prediktívna hĺbková analýza údajov, možno použiť na klasifikáciu, predikciu, odhadovanie a riešenie iných, podobných úloh. Informácie, ktoré poskytuje opisná hĺbková analýza údajov, slúžia na pochopenie skúmaného systému prostredníctvom objavenia vzorov a vzťahov v rozsiahlych množinách údajov.

Niektorí autori uvádzajú ďalší možný cieľ hĺbkovej analýzy údajov: *hľadanie nuggetov*. Pri hľadaní nuggetov sa požadujú nové, prekvapivé poznatky, ktoré nemusia pokrývať daný kontext ([2], s. 19).

4 Úlohy hĺbkovej analýzy údajov

Realizácia uvedených cieľov predikcie a opisu vedie k riešeniu týchto úloh hĺbkovej analýzy údajov ([6], s. 2):

- *klasifikácia* – objavenie prediktívnej učiacej funkcie, ktorá klasifikuje prípady do jednej z viacerých preddefinovaných tried,
- *regresia* – objavenie prediktívnej učiacej funkcie, ktorá priradzuje prípadom reálnu hodnotu prognózovanej (cieľovej) premennej,

³ Namiesto pojmu prípad sa v rovnakom význame používajú pojmy príklady alebo objekty. Z formálno-matematickej stránky prípad je n -rozmerný vektor údajov.

- *zhlukovanie* – opisná úloha, ktorá spočíva vo vytvorení konečnej množiny zhlukov prípadov,
- *sumarizácia* – opisná úloha, ktorá spočíva v nájdení kompaktného opisu množiny alebo podmnožiny prípadov,
- *modelovanie závislosti* – nájdenie lokálneho modelu, ktorý opisuje významné závislosti medzi premennými alebo hodnotami znaku v celej množine údajov alebo v jej časti,
- *detekcia zmien a odchýlok* – objavenie najvýznamnejších zmien v množine údajov.

Aj v klasifikácii úloh hĺbkovej analýzy údajov sa autori líšia. Niektorí uvádzajú tieto typy úloh: opis, odhadovanie, predikcia, klasifikácia, zhlukovanie, asociácia. Podrobnejšie v ([7], s. 11–17), ([3], s. 8–12).

Uvedieme ešte jednu definíciu hĺbkovej analýzy údajov: *Hĺbková analýza údajov je proces získavania rozličných modelov, súhrnov a odvodených hodnôt z danej množiny údajov* ([6], s. 5).

5 Aplikačné oblasti hĺbkovej analýzy údajov

Riešenie úloh hĺbkovej analýzy údajov možno nájsť v mnohých aplikačných oblastiach, napríklad pri:

- predikcii spotreby energetických médií,
 - predikcii kurzových pohybov na akciových trhoch,
 - klasifikácii zákazníkov bánk a poisťovní,
 - analýze dôvodov zmeny poskytovateľa služieb,
 - analýze príčin porúch rozličných strojov a zariadení,
 - analýze údajov o pacientoch v nemocniciach,
 - analýze nákupného koša
- a podobne.

Úlohy hĺbkovej analýzy údajov možno klasifikovať napríklad aj podľa vecného zamerania štúdie. V členení podľa tohto kritéria možno za najčastejšie úlohy hĺbkovej analýzy údajov považovať tieto úlohy.

Analyzovanie profilov. Ide o poznávanie súčasných a potenciálnych zákazníkov firmy. Dôkladná analýza profilov je často užitočná pri celení marketingových kampaní a vývoji nových produktov. Analýza profilov zahŕňa aj demografické faktory (vek, pohlavie, rodinný stav a pod.) a iné veličiny, dané potrebami firmy (dĺžka vzťahu, miera rizika, priemerné tržby a pod.). Profily sú užitočné, ak sú aplikované vnútri segmentov populácie.

Segmentácia zákazníkov organizácie sa realizuje pomocou zhlukovej analýzy na základe ich ziskovosti a trhového potenciálu. Cieľom je dobre poznať a porozumieť súčasným zákazníkom. Aplikáciou analýzy profilov na segmenty zákazníkov získava organizácia efektívny nástroj na hodnotenie akcií marketingu, alebo na riadenie rizík.

Modelovanie ozvy – cieľom je predpovedať, kto bude reagovať na ponuku produktu. To môže viesť k zníženiu výdavkov na marketingovú kampaň a k zvýšeniu jej účinnosti.

Modelovanie rizika sa využíva v tých odvetviach, kde organizácie prijímajú určité riziko strát pri ponuke svojich produktov, napr. v bankovníctve a v poisťovníctve. Cieľom je predpovedať pravdepodobnosť, že napríklad zákazník banky nebude schopný dodržiavať splátkový kalendár alebo zákazníkovi poisťovne sa stane poisťná udalosť. Do tejto skupiny možno zaradiť aj *modely na detekciu podvodov*.

Modelovanie aktivácie sa využíva na tvorbu modelov, ktoré slúžia na predpovedanie, či sa zákazník stane plnohodnotným zákazníkom. Využíva sa hlavne v oblasti finančných služieb. Napríklad zákazník s kreditnou kartou sa stane plnohodnotným zákazníkom vtedy, keď odpovie na ponuku, založí si účet a bude ho aj používať.

Modelovanie krížového a následného predaja – výsledkom sú modely na predpovedanie pravdepodobnosti, či si súčasný zákazník kúpi aj iné produkty od firmy (krížový predaj), alebo či bude zvyšovať objem nákupu produktu, ktorý od firmy nakupuje v súčasnosti (následný predaj).

Modelovanie odlevu zákazníkov má za úlohu odhaliť zákazníkov, ktorí sa chystajú prejsť ku konkurencii. Cieľom je identifikovať správanie sa zákazníkov, ktoré vedie k znižovaniu a k ukončeniu používania produktov firmy, aby firma mohla včas urobiť potrebné opatrenia.

Modelovanie čistej súčasnej hodnoty (Net Present Value) je určené na tvorbu modelov na prognózovanie ziskovosti produktu v stanovenom období. Výsledky sa prepočítajú na čistú súčasnú hodnotu toku peňazí v čase.⁴

Modelovanie hodnoty počas obdobia existencie (Life Time Value) má za úlohu vytvoriť modely na prognózovanie celkovej ziskovosti zákazníka za vopred stanovené obdobie.

Uvedený zoznam úloh určite nie je vyčerpávajúci. Ide o úlohy, ktoré sa v súčasnosti, v praxi, riešia najčastejšie. Objavili sa napríklad aj aplikácie z makroekonomickej oblasti.⁵

6 Korene hĺbkovej analýzy údajov

Hĺbková analýza údajov čerpá z viacerých disciplín, najviac však zo štatistiky a z teórie strojového učenia.

Štatistika zdôrazňuje matematickú presnosť, tvorbu teoretických základov predtým, ako sa budú na nich založené aplikácie overovať v praxi.

Strojové učenie vychádza väčšinou z počítačovej praxe. To vedie k praktickej orientácii, k snahe hľadať efektívne postupy bez toho, aby sa predtým hľadali for-

⁴ Postup výpočtu čistej súčasnej hodnoty toku peňazí v čase pozri napríklad v [11].

⁵ L. Bordoni and S. Spadaro: The Data Mining Techniques in the Macroeconomic Field, <http://www.actapress.com/Abstract.aspx?paperId=14080>.

málne dôkazy ich efektívnosti. Z toho vyplýva, že štatistika zdôrazňuje skôr pojem model, strojové učenie skôr pojem algoritmus.

Základné princípy modelovania v hĺbkovej analýze údajov majú korene v teórii riadenia.⁶

V prediktívnej hĺbkovej analýze údajov treba určiť matematický model neznámeho systému na základe množiny údajov. Problém určenia matematického modelu neznámeho cieľového systému na základe dvojíc údajov vstup – výstup sa nazýva *identifikácia systému*.

Identifikácia systému sa všeobecne realizuje v dvoch krokoch:

1. *Identifikácia štruktúry*. V tomto kroku sa aplikujú apriórne znalosti o cieľovom systéme na určenie triedy modelov, z ktorej sa vyberie najvhodnejší. Túto triedu modelov možno všeobecne označiť funkciou $y^* = f(\mathbf{x}, \mathbf{t})$, kde y^* je výstup modelu, \mathbf{x} je vektor vstupov a \mathbf{t} je vektor parametrov. Určenie funkcie f všeobecne závisí od riešeného problému a od skúseností a intuície riešiteľa.

2. *Identifikácia parametrov*. V tomto kroku sa aplikujú optimalizačné techniky na určenie vektora parametrov \mathbf{t} tak, aby výsledný model $y^* = f(\mathbf{x}, \mathbf{t})$ čo najlepšie opísal cieľový systém.

Obyčajne je na nájdenie vhodného modelu nevyhnutné viackrát zopakovať prvý aj druhý krok.

Celý postup možno opísať ako iteratívny proces s týmito krokmi:

1. Špecifikovať a parametrizovať triedu matematických modelov $y^* = f(\mathbf{x}, \mathbf{t})$, ktoré reprezentujú cieľový systém.

2. Hľadať parametre, ktoré najlepšie aproximujú údaje (pre ktoré je rozdiel $(y - y^*)$, kde y je výstup systému, minimálny).

3. Realizovať validačné testy, ktoré ukážu, či má identifikovaný model korektnú ozvu na neznámu množinu údajov (testovacia alebo validačná množina údajov). Keď sú výsledky validačných testov uspokojivé, ukončiť proces, keď nie, vrátiť sa k 1. kroku.

Keď nemáme apriórne informácie o cieľovom systéme, je identifikácia štruktúry systému veľmi ťažká a nezostáva nič iné, len ju realizovať metódou pokusov a omylov. V mnohých inžinierskych systémoch a priemyselných procesoch vieme často veľa o ich štruktúre. Štruktúra väčšiny cieľových systémov, na ktoré sa aplikujú techniky hĺbkovej analýzy údajov, je však obyčajne úplne neznáma alebo taká zložitá, že je nemožné získať adekvátny matematický model. Boli vytvorené nové techniky na identifikáciu parametrov, ktoré sú časťou množiny techník hĺbkovej analýzy údajov.

V hĺbkovej analýze údajov sa pojmom *model* obyčajne rozumie štruktúra, ktorá modeluje vzťahy medzi všetkými prípadmi. Pojmom *vzor* sa rozumie lokálny model, ktorý modeluje vzťahy medzi časťou prípadov.

⁶ Aplikuje sa hlavne v inžinierskych systémoch a v priemyselných procesoch.

7 Proces hĺbkovej analýzy údajov

Všimnime si proces hĺbkovej analýzy údajov podrobnejšie. Všeobecne sa odporúča postupovať takto ([6], s. 6–8):

1. *Špecifikácia problému a príslušných hypotéz.* Väčšina štúdií sa realizuje v určitej aplikačnej oblasti. Na vhodnú špecifikáciu problému sú obyčajne nevyhnutné poznatky a skúsenosti, ktoré sa týkajú špecifik tejto oblasti. Preto je veľmi užitočná spolupráca experta na techniky hĺbkovej analýzy údajov a experta na danú aplikačnú oblasť. V riešiteľskom tíme, samozrejme, nemôže chýbať odborník na údaje.

2. *Zhromaždenie údajov.* Po špecifikácii problému je nevyhnutné získať všetky dostupné údaje, ktoré možno použiť pri riešení problému. Niekedy je potrebné pracovať s údajmi, ktoré sú archivované v údajových súboroch, nie v databázach, niekedy s údajmi, ktoré sú uložené v rozličných systémoch, niekedy môže byť užitočné uvažovať aj o externých údajoch. Z hľadiska štatistiky, po zhromaždení údajov nemáme väčšinou žiadnu predstavu o ich rozdelení pravdepodobnosti, prípadne máme istú predstavu, implicitne danú charakterom procedúry zhromažďovania údajov. Je však veľmi dôležité, aby sme si boli istí, že údaje použité na odhadovanie modelu sú z toho istého základného súboru ako údaje použité na testovanie a aplikáciu modelu.

3. *Predspracovanie údajov.* V tomto kroku sa údaje získané na riešenie určitého problému pripravujú do formy, ktorá sa vyžaduje pri aplikácii jednotlivých metód. Tu sa realizuje aj identifikácia a prípadné odstránenie odľahlých údajov, prípadné doplnenie chýbajúcich hodnôt, kódovanie, výber znakov a ďalšie potrebné operácie s údajmi.

4. *Odhadovanie modelu.* Hlavnou úlohou v tomto kroku je výber a implementácia vhodných techník hĺbkovej analýzy údajov. Obyčajne sa implementuje viacero modelov. Doplnkovou úlohou je potom výber najlepšieho z nich.

5. *Interpretácia výsledkov a formulácia záverov.* Modely získané hĺbkovou analýzou údajov by mali pomáhať pri rozhodovaní. Často existuje konflikt medzi presnosťou modelu a medzi možnosťami jeho jednoduchej interpretácie. Zložitejší model poskytuje často presnejšie výsledky, je ale obyčajne ťažšie objasniť princípy jeho fungovania iným osobám, čo môže niekedy sťažovať jeho implementáciu. Rozhodovateľ sa môže zdráhať oprieť svoje rozhodnutia o výsledky generované jemu nepochopiteľným modelom. Problém interpretácie zložitých modelov treba považovať za osobitnú úlohu. Analytická správa, ktorá sa poskytne zákazníkovi, by nemala obsahovať stovky strán numerických výsledkov. Zákazník musí dostať stručné, obsahové informácie a jasné závery.

Treba zdôrazniť, že proces hĺbkovej analýzy údajov treba skutočne dôsledne chápať ako interaktívny a iteratívny. Hlboké pochopenie celého procesu je nevyhnutným predpokladom úspešných aplikácií.

Keď sme napríklad v kroku zhromaždenia a predspracovania údajov nepostupovali korektne, alebo sme v prvom kroku nevhodne špecifikovali problém, nemôžeme očakávať dobré výsledky, aj keď v štvrtom kroku aplikujeme „silné“ metódy na odhadovanie modelu.

Výsledky a závery, získané v štvrtom a piatom kroku, môžu napríklad indikovať potrebu vrátiť sa k prvému kroku a preformulovať problém a podobne.

8 Metodiky hĺbkovej analýzy údajov

Postupne začali vznikať metodiky, ktoré poskytujú používateľom určitý jednotný rámec na využitie hĺbkovej analýzy údajov pri riešení rozličných problémov. V týchto metodikách sú zovšeobecnené skúsenosti z úspešných projektov. Niektoré boli vypracované producentmi programových systémov⁷, iné vznikli mimo týchto producentov⁸.

8.1 Metodika SEMMA

Softvérový produkt firmy SAS, *Enterprise Miner*, má vlastnú metodiku hĺbkovej analýzy údajov. Jej názov SEMMA charakterizuje jednotlivé kroky postupu, odporúčaného pri aplikácii hĺbkovej analýzy údajov:

- *Sample* – tvorba vzorky (vyberanie),
- *Explore* – vizuálne preskúmanie a redukcia údajov,
- *Modify* – zoskupovanie prípadov a hodnôt znakov, transformácie údajov,
- *Model* – analýza údajov s využitím techník hĺbkovej analýzy údajov,
- *Assess* – porovnanie modelov a interpretácia.

Dôraz sa kladie hlavne na jednoduchú interpretáciu výstupov.

8.2 Metodika 5A

Túto metodiku ponúkla firma SPSS. Aj tu názov metodiky naznačuje odporúčané kroky postupu pri hĺbkovej analýze údajov:

- *Assess* – analýza potrieb projektu,
- *Access* – zhromaždenie potrebných údajov,
- *Analyze* – realizácia analýz s využitím techník hĺbkovej analýzy údajov,
- *Act* – transformácia poznatkov na akčné poznatky,
- *Automate* – implementácia výsledkov analýzy v praxi.

Podľa tejto metodiky, v prvom kroku je nevyhnutné stanoviť ciele, stratégie a procesy. Konkrétne by sa mali určiť údaje, ktoré treba zhromaždiť, mala by sa realizovať príprava na realizáciu projektov, v ktorej sa zabezpečia potrebné analytické nástroje a vzdelávanie a tréning zainteresovaných pracovníkov.

V druhom kroku sa zhromažďujú a pripravujú údaje. Získavajú sa vhodné údaje z údajových skladov a z iných interných zdrojov. Často je vhodné využiť aj externé zdroje údajov.

⁷ Napríklad metodika SEMMA firmy SAS alebo metodika 5A firmy SPSS.

⁸ Napríklad metodika CRISP-DM.

V treťom kroku sa realizuje aplikácia techník hĺbkovej analýzy údajov. V tomto kroku sa z údajov získavajú poznatky. Tu firma SPSS odporúča použiť viacero metód a porovnať ich výsledky a vhodnosť, aby sa čo najjednoduchšie a najrýchlejšie získalo najlepšie riešenie.

V štvrtom kroku by sa mali formulovať odporúčania a prijaté rozhodnutia. Tu sa získané poznatky menia na akčné poznatky. Výsledky by sa mali ponúkať v jasnej a zrozumiteľnej podobe.

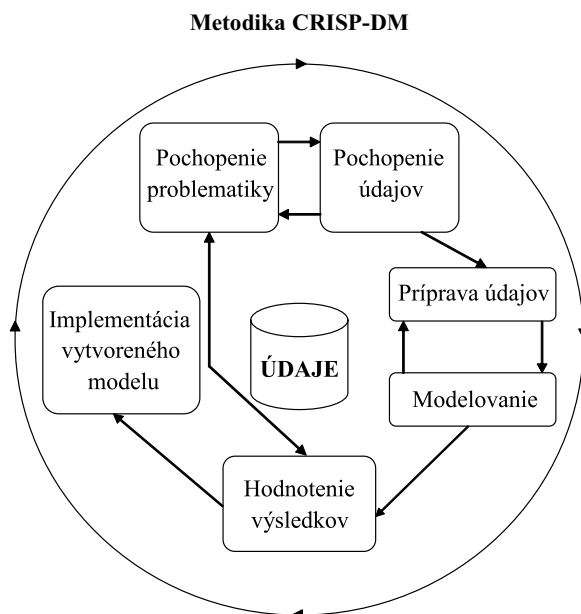
Piatym krokom je implementácia výsledkov v praxi. Tento krok obsahuje činnosti, ktoré súvisia s realizáciou prijatých rozhodnutí. Tu firma SPSS odporúča automatizovať analýzy tak, aby opakované úlohy nezaberali čas a aby bolo možné modely ľahko aktualizovať na základe nových výsledkov.

8.3 Metodika CRISP-DM

Metodika CRISP-DM⁹ vznikla v rámci Európskeho výskumného projektu. Cieľom bol návrh univerzálnej metodiky, použiteľnej v najrôznejších oblastiach.

Táto metodika je znázornená na obrázku č. 1. Podľa tejto metodiky tvorí životný cyklus projektu hĺbkovej analýzy údajov šesť fáz. Poradie týchto fáz nie je fixované. Výsledok, ktorý sa dosiahol v jednej fáze, ovplyvňuje výber nasledujúcich krokov. Často je potrebný návrat k niektorým krokom a fázam. Kružnica na obrázku symbolizuje cyklický charakter procesu hĺbkovej analýzy údajov.

Obr. č. 1



⁹ Cross-Industry Standard Process for Data Mining.

Záver

V článku sme sa pokúsili dať základnú charakteristiku hĺbkovej analýzy údajov spolu s možnosťami jej využívania v ekonomike. Ide o pomerne mladú disciplínu, ktorá sa však v posledných rokoch veľmi rýchlo rozvíja. Dôvodom je hlavne mimoriadna užitočnosť jej aplikácií.

Uvedli sme ciele, úlohy a aplikačné oblasti hĺbkovej analýzy údajov. Uvedené aplikačné oblasti určite neposkytujú vyčerpávajúci prehľad. Všeobecne sa prijíma predpoklad, že všade tam, kde sú k dispozícii databázy údajov, možno využiť metódy a modely hĺbkovej analýzy údajov.

V koreňoch hĺbkovej analýzy údajov poukazujeme na jej interdisciplinárny charakter. Čerpá hlavne zo štatistiky a z teórie strojového učenia. Využíva však aj niektoré poznatky o databázach, o umelej inteligencii a o neurónových sieťach.

Proces hĺbkovej analýzy údajov poskytuje všeobecné poznatky o postupe pri aplikáciách.

Nakoniec podávame stručný prehľad známych metodík hĺbkovej analýzy údajov. Nazdávame sa, že aplikácie hĺbkovej analýzy údajov budú mať aj u nás významnú úlohu pri urýchlení budovania znalostnej ekonomiky.

Literatúra

- [1] ADRIAANS, P. – ZANTINGE, D.: *Data Mining*. Harlow: Addison-Wesley, 1996. ISBN 0-201-40380-3.
- [2] BERKA, P.: *Dobývání znalostí z databází*. Praha: Academia, 2003. ISBN 80-200-1062-9.
- [3] BERRY, M. J. A. – LINOFF, G. S.: *Data Mining Techniques*. USA: J. Wiley and Sons, 2004. ISBN 978-0-471-47064-9.
- [4] FAYYAD, U. M. – PIATETSKY-SHAPIRO, G. – SMYTH, P. – UTHURUSAMY, R.: *Advances in Knowledge Discovery and Data Mining*. Cambridge MA: AAAI Press/MIT Press, 1996.
- [5] GUIDICI, P.: *Applied Data Mining. Statistical Methods for Business and Industry*. USA: J. Wiley and Sons, 2004. ISBN 0-470-84678-X.
- [6] KANTARDZIC, M.: *Data Mining. Concepts, Models, Methods, and Algorithms*. USA: J. Wiley and Sons, 2003. ISBN 0-471-22852-4.
- [7] LAROSE, D. T.: *Discovering Knowledge in Data. An Introduction to Data Mining*. USA: J. Wiley and Sons 2005. ISBN 978-0-471-66657-8.
- [8] LAROSE, D. T.: *Data Mining. Methods and Models*. USA: J. Wiley and Sons 2006. ISBN 0-471-66656-4.
- [9] PARALIČ, J.: *Objavovanie znalostí v databázach*. Košice: Elfa, 2003. ISBN 80-89066-60-7.
- [10] STANKOVIČOVÁ, I.: *Využitie štatistických metód v hĺbkovej analýze dát*. Dizertačná práca, FHI EU v Bratislave, 2005.
- [11] TEREK, M.: *Analýza rozhodovania*. Bratislava: IURA EDITION, 2007. ISBN 978-80-8078-131-6.
- [12] UHRINČAĎOVÁ, E.: Objavovanie poznatkov v databázach. *Ekonomika a informatika*, IV, 2006, č. 1.
- [13] UHRINČAĎOVÁ, E.: *Aplikácia objavovania poznatkov v databázach v oblasti analýz poľnohospodárskych fariem*. Dizertačná práca, FHI EU v Bratislave, 2008.
- [14] UHRINČAĎOVÁ, E.: Segmentácia poľnohospodárskych štruktúr vzhľadom na parametre produkčných systémov pomocou samoorganizujúcich sa máp. In: *Ekonomika poľnohospodárstva*, VIII, 2008, č. 4.