

Každá firma, která provozuje informační systémy, vlastní více či méně rozsáhlou databázi údajů, které obvykle ne zcela využívá. data mining je proces získávání znalostí, vztahů a dříve neznámých informací z dostupných dat. Pod tímto pojmem se skrývá mnoho různorodých technik a postupů, jimiž jsou vstupní „surová“ data upravována, čištěna a analyzována. Díky nesporným přínosům data miningu, zejména pro obchodní procesy, roste obliba tohoto druhu úloh a četnost jejich nasazení.

Informace získané prostředky data miningu lze začlenit do stávající IT infrastruktury a zpřístupnit je podle potřeby pracovníkům na všech úrovních organizační struktury společnosti (například managementu, pracovníkům call centra, pracovníkům, kteří vyřizují žádosti o úvěr apod.).

Data mining

Přeměna dat v hodnotné informace

Pavel Pilář

Ve stručnosti zmíníme alespoň několik nejvýznamnějších oblastí, kde se data mining s úspěchem využívá.

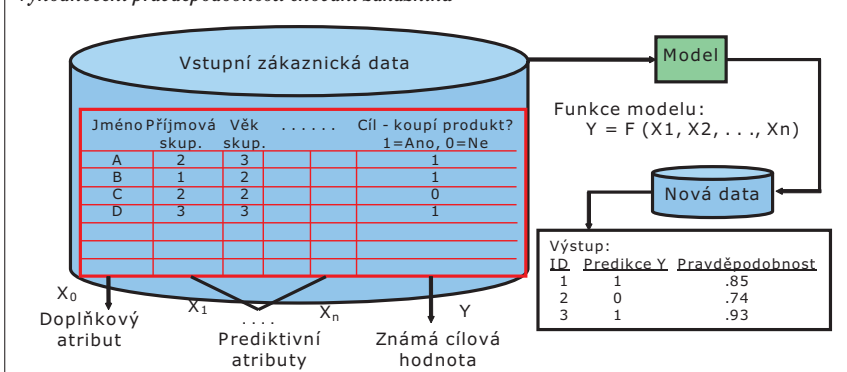
Marketing

Odhad pravděpodobnosti daného chování zákazníka. Pro optimální zacílení přímé reklamní kampaně je nesmírně cenná znalost pravděpodobnosti, s níž daný zákazník

zareaguje pozitivně na konkrétní nabídku. Metodami data miningu lze tuto pravděpodobnost odhadnout na základě dostupných dat o průběhu předchozích srovnatelných kampaní. Se znalostí této informace lze snížit náklady na cílenou reklamní kampaň o desítky procent při zachování její účinnosti. **Segmentace zákazníků.** V řadě velkých firem platí, že 90 % obrátu firmy tvoří 10 %

nejvýznamnějších zákazníků. Tuto základní formu segmentace na základě vybraných parametrů (v tomto případě výše obrátu) umožňuje už většina analytických nástrojů. Segmentace metodami data miningu umožní nalézt a identifikovat segmenty napříč desítkami atributů zákaznické databáze. Marketingová práce s jednotlivými dílčími segmenty je pak podstatně efektivnější než práce s kompletní zákaznickou databází. **Analýza nákupního koše.** Identifikace spotřebního chování. Z dat o prodejkách, jako jsou záznamy z prodejních terminálů nebo web logy internetových obchodů lze odvodit informace o nákupním chování zákazníků. Znalost těchto informací napomáhá efektivněji uspořádat obchodní katalog – ať už fyzický, nebo elektronický.

Vyhodnocení pravděpodobnosti chování zákazníků



Churn management

Odhalení potenciálního přechodu zákazníka ke konkurenci. Na základě analýzy dřívějšího chování zákazníků, kteří v minulosti přešli ke konkurenci, lze odvodit modely chování, které přechodu ke konkurenci předcházejí. Náklady na udržení stávajícího zákazníka jsou podstatně nižší než náklady na získání zákazníka nového. Na zákazníka, u něhož jsou identifikovány tyto typické vzory chování, lze pak cíleně působit tak, aby se potenciálnímu přechodu ke konkurenci předešlo.

Risk management, fraud detection

Odhad kreditního rizika. Odhad rizika opoždění plateb. Analýzou historických dat lze vytvořit modely, které identifikují potenciálně rizikový úvěr a míru tohoto rizika na základě demografických údajů a dalších dostupných informací.

Identifikace „potenciálního podvodníka“. Na základě analýzy chování zákazníků, u nichž byl zjištěn podvod nebo pokus o něj, lze odvodit modely pro chování takového zákazníka. Aplikací vytvořeného modelu na dostupná data lze identifikovat potenciální podvod.

Řízení výroby, management jakosti

Analýzou dat z průběhu automatizované výroby a jejich odchylek lze identifikovat problematické výrobní kroky, ať už z hlediska jakosti nebo z hlediska dodržení tempa výroby. Na základě takto zjištěných informací lze například do výrobního procesu doplnit dodatečnou kontrolu, která již v průběhu výroby odhalí rozpracované výrobky, které by po dokončení neprošly výstupní kontrolou.

Proces data miningu

Proces data miningu, jak je v současné době využíván většinou firem, lze rozdělit do následujících fází:

- **Definice problému** – prvním krokem v procesu je jasná definice problému z obchodního hlediska. Úspěšný data mining je vždy zahájen dobře definovaným projektem, musí být zřejmý zadavatel úlohy a její cíl.
- **Redefinice cíle v terminologii dostupných dat** – je třeba určit datové zdroje pro data nutná ke splnění zadání.
- **Příprava dat** – je obvykle nejzdoluhavější částí každého dataminingového projektu. Tato fáze v sobě zahrnuje konsolidaci dat z datových zdrojů, čištění dat a jejich přípravu do struktury vhodné pro data mining. Tato fáze je nesmírně důležitá pro celkový úspěch projektu – výsledné modely jsou tak dobré, jak dobrá jsou data použita pro jejich vytvoření.
- **Volba a tvorba modelu** – tato fáze zahrnuje volbu a následné využití převážně statistických nástrojů pro vytvoření a parametrizaci matematických modelů. Použitý model se volí dle cíle dataminingového projektu. Na základě vstupních dat se optimalizují parametry tak, aby model co nejlépe predikoval chování u neznámých dat.

- **Aplikace modelu** – proces, kdy se vybrané modely implementují do IT struktury firmy a vnitropodnikových procesů, aby mohly být využity ke zlepšení obchodních výsledků.

- **Interpretace a využití výsledků** – tato fáze zahrnuje využití zprovozněných modelů pro dosažení cíle definovaného při zadání projektu.

Začlenění data miningu do IT struktury

Postup začlenění dataminingového řešení do informační struktury podniku je závislý na dodavateli konkrétní technologie. V každém případě lze říci, že je vhodnější provádět data mining nad existujícím datovým skladem. Je-li data mining prováděn nad daty v primárních systémech, výrazně se tím komplikuje a prodlužuje fáze přípravy a čištění dat.

Technická architektura pro data mining je specifická pro jednotlivé dodavatele dataminingových řešení. Lze rozlišit dva základní modely architektury:

- **Oddělený dataminingový server** – v tomto uspořádání je dataminingový server fyzicky i technologicky oddělen od ostatních serverů společnosti. Před vlastním zpracováním dat na dataminingovém serveru je třeba všechna zpracovávaná data načíst na tento server.

- **Integrovaný dataminingový server** – v tomto uspořádání je dataminingový server technologicky integrován do jednoho prostředí s datovým skladem, serverem pro OLAP analýzy a případně i s částí primárních systémů. Zjednodušuje se tím správa dataminingového serveru i zajištění načítacích procesů. Tato varianta ale poněkud zužuje výběr možných dodavatelů řešení.

Prezentační vrstva dataminingového řešení obvykle zahrnuje různé varianty doručení informací – produktů data miningu cílovému uživateli. Zobrazení těchto informací

může být začleněno do provozních systémů. Uživatel má tak kromě informací o dřívějším nákupním chování zákazníka k dispozici například i predikci jeho dalšího nákupního chování. Jinou formou prezentace mohou být pravidelné reporty, například potenciálně problematických smluv či zákazníků. Další využití těchto informací je již dáno standardními vnitřními procesy.

Současnost a budoucnost data miningu

Ze současného stavu a vývoje dataminingových řešení lze vysledovat několik významných trendů do budoucna.

Rozšiřování. Stejně jako u datových skladů v nedávné době je patrné neustálé rozšiřování skupiny společností, které používají dataminingové postupy. Za tím stojí jednoznačně fakt, že efektivní využití získaných dat představuje pro firmy nemalou konkurenční výhodu.

Průběžně také zlevňují technické prostředky pro data mining. Lze tedy očekávat, že si dataminingová řešení postupně budou nacházet cestu do stále menších společností.

Prohlubování. Je obvyklé vyzkoušet technologii pro data mining v rámci firmy na jednodušším, pilotním projektu. U firem, které již dataminingová řešení zavedly, se obvykle řešení rozšiřuje o další oblasti.

Objevování nových směrů. Po prosazení data miningu v marketingu, bankovníctví, telekomunikacích a dalších oborech lze pozorovat postupné pronikání dataminingových technologií i do dalších oblastí. Za zmínku zde stojí například využití v medicíně, biochemii, genetice a jiných oborech.

*e-mail: system@ccb.cz
www.SystemOnLine.cz*

Autor článku působí ve společnosti Sophia Solutions.

— Inzerce —



s o p h i a sophisticated & advanced...

Sophia Solutions, s.r.o.

Doudlebská 1699/5, 140 00 Praha 4
tel.: +420 226 517 800, fax: +420 226 517 801
e-mail: info@sophias.cz, <http://www.sophias.cz>

Váš Business Intelligence partner

Plánovací, controllingové a modelovací systémy, datové sklady, manažerské informační systémy, řešení pro podporu rozhodování