

ROZHOVOR

Objevte poklady ukryté ve firemních datech

Možná ani netušíte, jaké poklady se skrývají ve vašich firemních datech. Záleží ale jen na dobře zvolené metodě data miningu a následné výstupy vám mohou přinést úsporu nákladů, říká mj. rozhovoru Miloš Uldrich, technický konzultant a analytik společnosti StatSoft.



Nejčastější dataminingovou úlohou je klasifikace. Pod tím si představme odhad pravděpodobnosti vzniku určité situace – porucha energetické soustavy, odchod klienta, pojištný podvod, kladná reakce na slevu, postoupení pohledávky apod.

Co všechno může firma očekávat od data miningu?

Cílů, resp. přínosů pokročilejšího modelování nad daty je víc. Očekávat můžeme především znalosti a nové poznatky o chování našich klientů, které pomocí standardních reportovacích nástrojů, jež obsahují pouze základní sumarizační techniky a vizualizaci, nikoliv pokročilé modelovací algoritmy, nezjistíme. Pokud se přehoupneme do oblasti průmyslu, výstupem jsou podrobné znalosti o chování systému výrobního nebo energetického.

V této oblasti například jeden z našich zákazníků odhaduje chování produktu na základě nepřetržitě generovaných dostupných dat z kontrolních čidel. Vytvořený dataminingový model neustále informuje o stavu produktu. Ztratí-li

produkt konzistenci, pracovník kontroly je informován pomocí odhadu z modelu a okamžitě provádí kroky k nápravě. Alternativou, kterou tento postup vytlačil na druhou kolej, je chemická analýza produktu, na jejíž výsledky se ale musí čekat velmi dlouho, takže ztráty jsou neúměrně větší.

V oblasti klientských dat v komerční sféře (finanční instituce, mobilní operátoři, slevové portály apod.) je jedním z cílů v podstatě to samé. Na základě dostupných, neustále generovaných zákaznických dat, tím myslím např. provolané minuty, bankovní transakce, průchody webem, realizované platby apod., odhadnou, že se s konkrétním klientem děje něco nestandardního. Typickou úlohou je stanovení pravděpodobnosti odchodu jednotlivých

zákazníků ke konkurenční společnosti (*churn prediction*), samozřejmě s co možná největší mírou automatizace. Pokud rozpoznám u konkrétních zákazníků tuto tendenci v předstihu, je podstatně větší šance na retenci těchto skupin. To je pouze jeden z mnoha příkladů. Další obvyklou úlohou je detekce podvodů (*fraud detection*), pokročilá zákaznická segmentace, která slouží marketingovému oddělení apod.

Jaké přínosy může mít data mining pro obchodní procesy a pro výrobní procesy?

Pokud je hlavním obchodním procesem například prodej pojištění nebo poskytnutí půjčky, na základě klientské historie mohou pomoci dataminingového modelu přesněji ▶

▶ odhadnout, kdo nebude schopen splácet, kterou pohledávku vymáhat standardním postupem a kterou ne, neboť standardní kroky nic nezpečou. Data mining tedy vstupuje do obchodního procesu jako jedna z činností, díky které získávám ohodnocení nějakého vstupu.

V oblasti propagace firemních produktů potom můžeme hovořit o zvýšení efektivity. Typické dataminingové úlohy pomáhají s vytipováním skupin klientů, které má smysl oslovovat s nabídkou, slevou, novým produktem. Samozřejmě je zde nějaká chybovost, nicméně výhody jsou značné, kromě úspory nákladů jde také o podporu pozitivního vnímání značky, neboť doprovodné služby nenabízíme automaticky každému.

U výrobních procesů je situace trochu jiná, zde jde především o predikci kvality produktu, odhad budoucí poruchy systému apod. Cílem je např. zabránit delší odstávce výroby. Společným znakem těchto činností je zpravidla jejich opakované použití nad stále novými daty.

Je možné díky těmto výstupům identifikovat obchodní příležitosti?

Vliv na nové příležitosti bude zprostředkovaný. Na B2C trhu mohou např. pomocí shlukové analýzy identifikovat určité homogenní zákaznické segmenty, tedy skupiny klientů, kteří jsou si podobní z hlediska určitých kombinací vlastností, a tím nemyslím pouze základní demografická hlediska a hodnotovou segmentaci, ale segmentaci napříč desítkami parametrů. Na základě takových výstupů potom přesněji vymezím svoje zákazníky a jejich potřeby a možnosti. Marketingové aktivity přizpůsobuji jednotlivým zákaznickým skupinám, celý proces je mnohem efektivnější.

Která oddělení ve firmě by měla mít přístup k výstupům?

Záleží na předmětu podnikání a na obsahu firemní databáze, ten určí možnosti pokročilejších analýz pro různá oddělení ve společnosti. V oblasti finančních služeb, pojišťovnictví ale i Telco jsou výstupy z těchto pokročilých analýz významným pomocníkem při tvorbě a plánování marketingových aktivit (*cross/up-sell*), další nepostradatelnou úlohou je modelování rizika,

jako např. opožděné platby, podvody, významné změny směnných kurzů atd. Tedy oddělení řízení rizik, marketingové oddělení, ale také finanční a samozřejmě klíčová opatření nad aktuálními zákaznickými daty musí mít vždy podporu managementu, ten v data miningu získá významného spojence pro podporu strategických rozhodnutí.

Do jaké míry je možné dataminingové metody upravit podle oboru podnikání firmy?

Do data miningu spadá široký soubor analytických metod. Každá metoda je vhodná pro určitý typ cíle, tedy účel analýzy. Cílem může být nalezení složitější zákaznické struktury nebo nalezení asociací mezi jednotlivými položkami na webu. Nejčastější dataminingovou úlohou je klasifikace. Pod tím si představme odhad pravděpodobnosti vzniku určité situace – porucha energetické soustavy, odchod klienta, pojistný podvod, kladná reakce na slevu, postoupení pohledávky apod.

Použitá metoda se napříč obory lišit vůbec nemusí, liší se data na vstupu. Výstupem, pokud zůstaneme u klasifikační úlohy, bude vždy nějaké skóre, tedy pravděpodobnost vzniku sledované situace.

Jaký je nejčastější důvod – proč se na vás zákazníci obracejí? Co chtějí „najít“?

Důvody jsou velmi různorodé. V oblasti webových aplikací (e-shop, slevové portály) hledají nové příležitosti pro stávající klienty, tedy efektivnější marketing pro vybrané skupiny. Finanční instituce obvykle mají jasně definované cíle těchto analýz, jde o již zmiňované různé typy klasifikačních úloh. V oblasti energetiky je důvodem nejčastěji riziko poruchy, resp. efektivnější předpověď poruchy systému na základě neustále generovaných provozních dat.

Jaké jsou základní ukazatele toho, že užívání přináší požadovaný efekt?

Opět záleží na typu úlohy. Z obecného hlediska jde především o zvýšení tržeb a úsporu nákladů. Jednotlivé segmenty zákazníků oslovuji s produktem, jenž reflektuje jejich konkrétní potřeby a zvyklosti, které na základě analýzy dat

výrazně zpřesnil. Dalším ukazatelem je již zmiňované omezení výdajů, a to v různých oblastech (retence zákazníka, vymáhání dluhů, efektivnější marketing atd.). Někdy je potřeba se na úspory dívat ze skutečně dlouhodobého měřítka, jindy je naopak efekt velmi znatelný již v krátké době po nasazení dataminingového modelu na data, jako například efektivnější identifikace pojistných podvodů, zabránění poruchy výrobní soustavy, kdy jakákoliv neplánovaná odstávka je neúměrně drahá.

Jaké kritické ukazatele podstatné pro risk management je možné monitorovat?

V oblasti průmyslu jde o ukazatele specifické pro danou produkci (energetika, výrobní linka atd.), které generují sledovaná zařízení a čidla v rámci celého výrobního systému. U finančních institucí máme několik základních typů rizik: tržní, kreditní, provozní apod. Pokud se podíváme například na riziko odchodu klienta, základním ukazatelem je jeho chování k produktům společnosti. Tedy například to, jak často používá bankomaty, jak často přeposílá peníze na jiný účet, změny v počtu využívaných produktů společnosti apod. Tyto ukazatele nazýváme behaviorální data, jde o informace s největší hodnotou pro pokročilejší odhady dalšího chování. Na základě těchto dat se potom snažíme zjistit, jestli ti, co odcházejí, nemají například podobné kombinace služeb apod.

Jak efektivně investovat do těchto systémů?

Zde je třeba trh opět rozdělit. Bankovní domy, které mají ohromné zákaznické databáze, se již dávno přesvědčily, že bez efektivního využití těchto dat se prakticky neobejdou. Podobná situace je u pojišťovnictví nebo Telco. Výhody využití transakčních dat si v posledních letech uvědomují také firemy podnikající výhradně na internetu, neboť i ony mají obrovské množství registrovaných uživatelů. V oblasti průmyslu se nedá jednoznačně mluvit o vhodnosti těchto metod, zde skutečně záleží na povaze produktu. Velký a zatím, alespoň dle mého názoru, nevyužitý potenciál je v oblasti energetiky.



Pokud se podíváme například na riziko odchodu klienta, základním ukazatelem je jeho chování k produktům společnosti. Tedy například to, jak často používá bankomaty, jak často přeposílá peníze na jiný účet, změny v počtu využívaných produktů společnosti apod. Tyto ukazatele nazýváme behaviorální data, jde o informace s největší hodnotou pro pokročilejší odhady dalšího chování.

Co zpravidla o investici do dataminingového nástroje rozhoduje a co by mělo rozhodovat?

Vzhledem k určité náročnosti nasazení dataminingového řešení by měla rozhodovat, a v praxi tomu tak skutečně je, konkrétní potřeba nebo požadavek, jenž vzešel např. z interního auditu určitých firemních procesů. Požadavkem se např. rozumí snížení počtu ztracených pohledávek, které byly puštěny do standardního procesu vymáhání, optimalizace pojistné sazby, eliminace většího počtu rizikových klientů, efektivnější cross-up/sell. Požadavek může být i obecnějšího rázu, například pochopení stávajících klientů nebo využití velkého množství nashromážděných dat.

Co je podstatné pro úspěšný data mining?

Kvalita vstupních dat. Konzistentní a dlouhodobě ukládaná zákaznická data jsou nutným předpokladem pro složitější typy analýz a pro tvorbu jakýchkoliv predikčních modelů chování. Firma musí ukládat co možná nejvíc informací o jednotlivých klientech, konzistentně a po nějakou dobu. Musí si vést záznamy, kteří klienti firmu opustili, kteří registrovaní uživatelé zareagovali na slevu, kteří nikoli, která pojistná plnění byla podvodná apod. Teprve potom, na základě těchto historických dat, mohou vytvářet zmiňované modely chování a s jejich pomocí s dobrou úspěšností odhadovat chování nově příchozích žadatelů o úvěr, pojistných plnění apod.

Můžete uvést nějaké zajímavosti, co se podařilo díky data miningu vypátrat?

Výstupy z analýz jsou know-how každé společnosti, přestože zákaz-

nické chování je v rámci konkrétní oblasti často velmi podobné, tedy alespoň v rámci geografického nebo demografického určení. Není výjimkou, že zpracováváme data, u kterých jsou položky (klienti) a identifikační atributy překódovány. Samozřejmě, pro určitý typ analýz je skutečně nutné znát pozadí dat, nastavená expertní pravidla v dané společnosti apod. Jen tak je možné docílit plného pochopení dat a z analýzy vzešlých vazeb. Pomocí data miningu se nám podařilo například zjistit, že se významně liší chování dlužníků při dluhu převyšujícím určitou dlužnou částku, dále se nám například podařilo identifikovat kombinace faktorů, které s velkou pravděpodobností určují podvodnou žádost o pojistné plnění.

Podle čeho vybírat dodavatele?

Při výběru dodavatele, v případě zavádění data miningu do společnosti, je velkou výhodou (díky značné flexibilitě úpravy řešení pro specifické a často se měnící potřeby společnosti), pokud je dodavatel zároveň i výrobcem implementovaného dataminingového nástroje. Finální softwarové řešení musí obsahovat všechny metody, které jsou nezbytné pro konkrétní požadavek. Pod tím stojí vlastní ovládání softwarového produktu, zde se nároky velmi odlišují podle oddělení, které bude software využívat. V dnešní době je klíčový také celkový přístup dodavatelské společnosti směrem k zákazníkovi, licenční politika, celkové náklady na řešení, tzv. *total cost of ownership* (TCO), snadná implementace do IT infrastruktury s využitím standardních technologií, dlouhodobá podpora uživatelům, možnosti jejich vzdělávání apod.

Jaké jsou trendy v této oblasti (např. nové obory, kde se data mining prosazuje)?

Trend vidím ve stále širším využití napříč všemi obory lidské činnosti, jako je IT bezpečnost, zdravotnictví, genetika, energetika a mnoho dalších. Dalším trendem je rozšiřování stávajícího použití na další firemní procesy, které dříve neměly „zrcadlo“ v podobě historických dat uložených v databázi a v neposlední řadě také zvýšená poptávka po outsourcingu a pilotních analýzách. •

kamila.fuchsova@gmail.com



ING. MILOŠ ULDRICH

Pracuje jako Technical Consultant/Analyst ve společnosti StatSoft CR. V letech 2003–2009 vystudoval obor Systémové inženýrství a informatika na Provozně ekonomické fakultě České zemědělské univerzity v Praze. Při studiu se zaměřoval zejména na oblast statistiky a modelování. Po ukončení magisterského studia nastoupil na plný úvazek do společnosti StatSoft jako technický konzultant a analytik, kde se specializuje zejména na oblast business analýzy a rozvoje příležitostí. V roce 2011 se stal kvalifikovaným lektorem odborných statistických kurzů.