
Segmentace klientské databáze pro CRM

N. Kaspříková

<http://data.tulipany.cz>

Úvod

Segmentace patří mezi základní nástroje marketingového řízení. Pro efektivní vykonávání činností, které lze zařadit do často diskutované oblasti řízení vztahů se zákazníky (CRM), ve společnostech s velkým počtem klientů přirozeně vzniká potřeba dostupnosti vhodné segmentace klientů. V situacích, kdy ještě neexistuje přesná představa o tom, které charakteristiky by měly být pro rozdělení klientů do segmentů určující, se přistupuje k analýze dat o klientech a jejich chování.

Segmentační modely sice nepřinášejí tak rychlé a dobře měřitelné výsledky jako predikční modely (zejména ty pro výpočet ukazatelů vyjadřujících sklon ke koupi určitého produktu, používané pro výběry optimálních cílových skupin pro nabídku produktu) a je jim tak nejspíše oprávněně věnována menší pozornost, jejich úloha je však obtížně zastupitelná. Podporují lepší poznání chování a potřeb klientů a umožňují lépe zaměřit marketingovou komunikaci. Nejsou-li pro některé produkty dostupné predikční modely pro výpočty propensity-to-buy skóre, lze provádět cílení kampaní alespoň podle segmentů. Segmenty též mohou posloužit jako podklad pro nejspíše správnou obecnou stratifikaci zvyšující korektnost při měření účinnosti některých marketingových akcí. Při predikčním modelování lze někdy dosáhnout lepších výsledků tvorbou několika jednodušších modelů přiřazených skupinám klientů podle segmentů než jediným modelem určeným pro celou klientelu. Pokud nelze splnit všechny požadavky kladené na segmentaci jedinou segmentací, lze při zvládnutí vyšších nároků na obsluhu modelů používat několik segmentací zaměřených na jednotlivé účely. Segmentační modely je také možné konstruovat jako hierarchické a například stanovit základní úroveň podle jednoduché apriorní definice a dále pokračovat segmentací behaviorální a demografickou vytvořenou na základě analýzy dat.

Velké společnosti obvykle disponují rozsáhlou datovou základnou, často datovým skladem, ve kterém jsou sledovány charakteristiky klientů a jejich chování ve vztahu ke společnosti. K požadavkům na možnost snadné interpretace segmentace a porozumění vlastnostem segmentů, podobnost klientů ze stejného segmentu, zvládnutelný počet segmentů a jejich stabilitu v čase přistupují nároky na efektivitu procesu tvorby modelu (objevování definic segmentů, tvorba pravidel pro určování příslušnosti klientů k segmentům) a rychlost jeho aplikace (přiřazování klientů k segmentům).

K tvorbě segmentačního modelu se nabízí klasický aparát shlukové analýzy (například [1], [2], [3]). Ale snad žádný postup není bez problémů vyhovující všem výše zmíněným požadavkům. Použití takových metod pro tvorbu modelu obvykle vyžaduje provedení některých v zásadě arbitrárních, bohužel však klíčových rozhodnutí, nad jejichž dopady uživatel nemá dostatečnou kontrolu. Poměrně citlivá je například už základní otázka volby používané vzdálenosti nebo volby vhodného informačního kritéria pro optimalizaci.

Řešení segmentační úlohy

V jedné z velkých finančních institucí vznikl požadavek na úpravu stávající segmentace ve smyslu dalšího rozčlenění jednoho ze segmentů pro fyzické osoby, který byl shledán jako nedostatečně stejnorodý pro účely marketingového řízení. Hlavní cíl je formulován obecně jako lepší poznání chování a potřeb klientů. Segmentace by tedy nejspíše měla zohlednit charakteristiky demografické i behaviorální.

Po přípravě dat, výpočtech odvozených proměnných (například některých ve zdrojových databázích ne přímo dostupných agregací produktů, zůstatků na účtech podle jednotlivých typů a počtů a objemů transakcí podle typů) a průzkumu dat se při úvodních experimentech se shlukováním metodou k-průměrů pro různé počty segmentů objevily některé přirozené (a očekávané) skupiny

(například studenti nebo neaktivní klienti), ovšem interpretace jiných segmentů byla obtížná. Navíc byly výsledky dosti proměnlivé podle zadaného počtu segmentů. Přesvědčivý výsledek nedalo ani použití algoritmu k-průměrů na překódovaná data, ve kterých byly hodnoty vstupních proměnných nahrazeny průměrnou hodnotou ročních výnosů ze vztahu s klientem platnou pro danou kategorii vstupní proměnné. Nebylo jasné, podle čeho dobře rozhodnout o přesném počtu segmentů. V této fázi bylo přistoupeno k analýze dat algoritmem tvorby rozhodovacího stromu (implementace algoritmu CART) s úrovní ročních výnosů jako cílovou proměnnou. Rozhodovací strom bývá vnímán především jako nástroj pro tvorbu predikčních modelů a případně pro průzkum dat, ale může být s výhodou použit i pro řešenou segmentační úlohu. Tento přístup vlastně představuje formu hierarchického shlukování a umožňuje uživateli následné ruční úpravy stromu nahrazení podstromu listem nebo prohození uzlů (rodiče s potomkem) pro zjednodušení modelu a dosažení srozumitelné a zároveň stále dostatečně přesné segmentace se zvládnutelným počtem segmentů (v reprezentaci rozhodovacího stromu listů) přiměřené velikosti.

Problematická může být volba cílové proměnné pro rozhodovací strom. Hypotéza, že roční výnosy generované vztahem s klientem souvisí s fází životního cyklu klienta, objemem depozit, typy a počtem vlastněných produktů a s mírou transakční aktivity klienta, se jeví jako věrohodná. Jde o proměnnou dostatečně navázanou na další sledované proměnné a vytváří tak dobrý základ pro segmentaci. Navíc i tato proměnná přímo je ve společnostech pochopitelně předmětem velkého zájmu a je přirozené, když se v segmentaci projevuje.

Výsledná pravidla pro přiřazení klientů do segmentů mohou vypadat přibližně takto:

Jestliže klient:

vlastní účet pro děti → Děti

...

jinak: vlastní úvěrový produkt → Úvěrovaní klienti

jinak: je vlastníkem konta pro studenty → Studenti

jinak: je klientem méně než 5 měsíců → Noví klienti

jinak: (měsíční poplatky < 31 Kč) → Spící klienti

jinak: (objem depozit > 52 000 Kč) → Aktivní s depozity

jinak: (měsíční poplatky > 110 Kč) → Aktivní

jinak: Středně aktivní.

Pravidla mají být aplikována postupně ve stanoveném pořadí, což například vyjadřuje obchodní rozhodnutí přistupovat ke studentům, kteří mají úvěr, spíše jako k úvěrovaným klientům než ke studentům bez úvěru. Podobně do technického segmentu, do kterého obecně patří noví klienti, kteří zatím neměli dobrou příležitost projevit ve společnosti své chování, nespádnou klienti, kteří se již projevili například jako (úspěšní) zájemci o úvěr. Pravidla jsou velmi dobře srozumitelná a poskytují rovnou i základní popis segmentů. Definice segmentů lze snadno převést do jazyka SQL využívaného v databázových systémech a použít v databázovém prostředí pro pravidelné přiřazování klientů do segmentů.

Závěr

Problematika řešená při segmentačních úlohách je poměrně široká a těžko může být podán její úplný přehled. Navrhovaný postup samozřejmě nemusí být bez úpravy přenositelný do odlišného datového a obchodního prostředí. Ukázalo se, že pro splnění požadavků na segmentaci je vhodné vytvořit rozhodovací strom s dobře zvolenou cílovou proměnnou (roční výnosy ze vztahu s klientem) a následně provést jeho úpravu (zjednodušení), resp. převod na sadu rozhodovacích pravidel. Při tomto postupu je snadná integrace apriorních představ a požadavků uživatelů a výsledků analýzy dat. Navíc získaný výstup lze snadno nasadit i v rozsáhlém databázovém prostředí a je uživateli dobře srozumitelný. Zároveň použití prostředku pro automatizované získání znalostí z dat vytváří dobré předpoklady k tomu, aby výsledný model vyhovoval i z hlediska rozlišitelnosti a stability segmentů.

Přehled (zne)užitých zdrojů:

- [1] HAVRÁNEK, T. a kol.: Matematika pro biologické a lékařské vědy. Academia. Praha, 1981.
- [2] ROUSSEEUW, P. - STRUYF, A. - HUBERT, M.: Clustering in an Object-Oriented Environment. Journal of Statistical Software, Volume 1, 1996, Issue 4.
- [3] CHEESEMAN, P. - STUTZ, J.: Bayesian Classification (AutoClass): Theory and Results. In Fayaad, U. - Piatetsky-Shapiro, G. - Smyth, P. - Uthurusamy, R. (eds.): Advances in Knowledge Discovery and Data Mining, AAAI Press, 1995.

Klíčová slova: segmentace, klasifikace, data mining, shluková analýza, customer relationship management.