

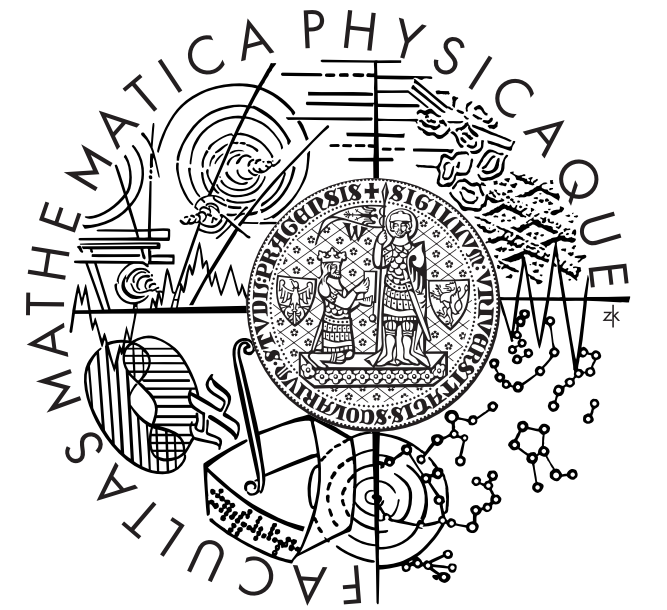


Statistika a spotřebitelské úvěry

David Legát

legat@karlin.mff.cuni.cz

Katedra matematické statistiky a pravděpodobnosti,
MFF UK, Praha



Shrnutí

V mnoha oblastech prodeje se využívají statistické modely pro předpovídání chování klientů. Nejinak je tomu v případě poskytování spotřebitelských úvěrů, kde tyto modely slouží jak pro řízení rizika v podobě velkého množství nesplacených úvěrů, tak pro různé marketingové kampaně. Tento poster obsahuje popis několika vybraných problémů, které se opakovaně vyskytují při vytváření požadovaných modelů, spolu s doporučeními, jak tyto problémy řešit.

Vytvoření modelu

Fáze vytváření

- Sběr a čištění dat
- Vytvoření modelu a odhad parametrů
- Implementace modelu do systému

Používání modelu

- Sledování výkonnosti výsledného modelu

Klasifikační úloha

- Logistická regrese
- Klasifikační stromy

Vybrané problémy při modelování

- Definice cílové proměnné
- Výběr vysvětlujících proměnných
- Transformace proměnných
- Úvaha o zamítnutých

Transformace proměnných

Často se používá logistická regrese

- Daný tvar závislosti odhadované pravděpodobnosti na prediktorech $p = 1/(1 + \gamma e^{-\beta x})$ (monotónní).

Možná řešení

- Transformace proměnných (Log, Box-Cox)
- Kategorizace proměnných
 - + Jednoduchá interpretace, řeší odlehlá pozorování
 - Vyšší počet odhadovaných parametrů

Váhy zřejmosti (weights of evidence)

- Pro každou hodnotu kategoriální proměnné vypočteme

$$WOE = \log(P_{good}/P_{bad})$$

$$P_{good} = \frac{\# \text{dobrých v dané skupině}}{\# \text{dobrých celkem}} \sim \text{Proc. dobrých}$$

$$P_{bad} = \frac{\# \text{špatných v dané skupině}}{\# \text{špatných celkem}} \sim \text{Proc. špatných}$$

Určení cílové proměnné

Účel

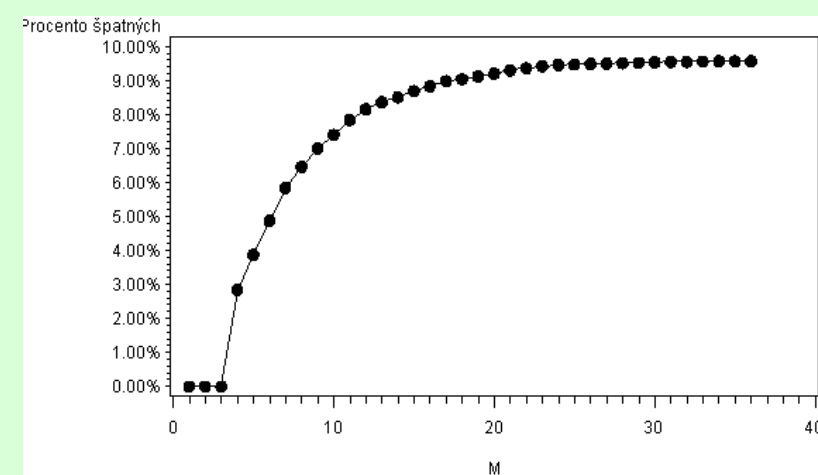
- Definovat pojem typu *špatná událost* (co to znamená, řekneme-li, že klient špatně splácí)
- Minimalizovat dobu potřebnou k určení špatné události.

Možné varianty definic pro riziková skóre

- Momentálně špatný: Klient je po M měsících po poskytnutí úvěru ve zpoždění se splácením alespoň o K měsíců.
- Někdy špatný: Klient se někdy v průběhu prvních M měsíců po poskytnutí úvěru dostal do zpoždění se splácením alespoň o K měsíců.

Volba parametrů

- Volba K ve spolupráci se specialisty na vymáhání.
- Volba M na základě závislosti počtu špatných úvěrů na M .



Úvaha o zamítnutých

Nově vytvořený model bude využíván na všechny nové žadatele o úvěr. U žadatelů, které jsme v minulosti zamítli, však nemáme informaci zda by spláceli dobře nebo ne.

Používané přístupy

- Náhodně vybrat některé žadatele, kterým úvěr poskytneme, přestože by neuspěli v normálním posuzovacím procesu.
- Vytvořit model pro výpočet pravděpodobnosti P_i přijetí úvěru. Následně použít váhy $w_i = 1/P_i$ při vytváření modelu.
- Nasimulování modelované události u zamítnutých úvěrů
 1. Vytvořit předběžný model za použití přijatých žádostí.
 2. Dosimulovat cílovou proměnnou u zamítnutých žádostí.
 3. Vytvořit finální model za použití všech žádostí.

Užitečné odkazy

Web:

- www.phil.frb.org/files/br/brso97lm.pdf
- www.creditinfocenter.com/credireports/scoring/
- www.consumersunion.org/finance/scorewc200.htm
- www.fairisaac.com; www.myfico.com
- www.nacm.org

Ostatní:

- McCullagh, P., Nelder, J., 1989. Generalized Linear Models, 2nd Edition. Chapman & Hall, London.
- Fahrmeir, L., Tutz, G., 1994. Multivariate Statistical Modelling Based on Generalized Linear Models. Springer, New York.
- Antoch, J., Gregoire, G., Jarušková, D., 2004. Detection of structural changes in generalized linear models. Statistics & Probability Letters 69, 315 - 332.

Výběr vysvětlujících proměnných

Účel

- Maximalizovat výkonnost modelu (Lift křivka, ROC křivka, Lorencova křivka).
- Zajistit stabilitu v čase.

Výběr prediktorů

- Vyloučení proměnných, které
 - nejsou dostatečně diskriminující
 - nemají stabilní chování v čase
- Různé varianty dopředných (forward) výběrových strategií
- Snaha o to, aby se v modelu vyskytovalo vyšší množství proměnných s malou prediktivní silou, než několik málo proměnných s vysokou prediktivní silou.