Regrese v modelech oprav

Petr Novák

KPMS MFF UK

11. září 2012

Modely oprav - o co jde

- Studujeme data o historii zařízení, které podléhá opotřebení.
- Když se porouchá, je nutné provést opravu.
- Poruchám se snažíme předcházet preventivní údržbou.
- $T_1, ..., T_n$ časy zásahů (oprav nebo údržeb).
- $\Delta_1, ..., \Delta_n$ indikátor zda v j-tém čase byla provedena oprava.
- $X(t)$ vysvětlující proměnná.

Modely oprav - o co jde

- Studujeme data o historii zařízení, které podléhá opotřebení.
- Když se porouchá, je nutné provést opravu.
- Poruchám se snažíme předcházet preventivní údržbou.
- $T_1, ..., T_n$ časy zásahů (oprav nebo údržeb).
- $\Delta_1, ..., \Delta_n$ indikátor zda v j-tém čase byla provedena oprava.
- $X(t)$ vysvětlující proměnná.

- Chceme rozumný popis vlivu oprav, údržby a regresorů na životnost.

Data jako čítací procesy

- Zavedeme čítací procesy oprav a údržeb

$$N_\bullet(t) = \sum_{j=1}^{n} I(T_j \leq t, \Delta_j = 1), \qquad M_\bullet(t) = \sum_{j=1}^{n} I(T_j \leq t, \Delta_j = 0).$$

- Označíme rizikovou funkci

$$\lambda(t) = \lim_{h \to 0} P(N_\bullet(t+h) - N_\bullet(t) \geq 1 | \mathcal{H}(t))/h$$

kde $\mathcal{H}(t)$ značí historii událostí do času t
a kumulativní rizikovou funkci $\Lambda(t) = \int_0^t \lambda(s) ds$.

Data jako čítací procesy

- Zavedeme čítací procesy oprav a údržeb

$$N_\bullet(t) = \sum_{j=1}^{n} I(T_j \leq t, \Delta_j = 1), \qquad M_\bullet(t) = \sum_{j=1}^{n} I(T_j \leq t, \Delta_j = 0).$$

- Označíme rizikovou funkci

$$\lambda(t) = \lim_{h \to 0} P(N_\bullet(t+h) - N_\bullet(t) \geq 1 | \mathcal{H}(t))/h$$

  kde $\mathcal{H}(t)$ značí historii událostí do času t
  a kumulativní rizikovou funkci $\Lambda(t) = \int_0^t \lambda(s) ds$.
- Předpokládejme, že oprava sice vrátí prvek jen do stavu těsně před poruchou, ale má vliv na rizikovou funkci.
- Rizikovou funkci vhodně parametrizujeme, chceme odhadnout parametry metodou max. věrohodnosti.
- Log-věrohodnost lze přepsat jako

$$l = \sum_{j=1}^{n} \Delta_j \log \lambda(T_j^-) - \int_0^{T_n} \lambda(t) dt.$$

Coxův model

- V Coxově modelu působí regresory multiplikativně na rizikovou funkci.
- Předpokládáme, že každá oprava či údržba multiplikativně sníží nebo zvýši riziko, stejně tak případné regresory. Uvažujeme rizikovou funkci ve tvaru (Percy & Alkali 2005):

$$\lambda(t) = \lambda_0(t) e^{M_\bullet(t)\rho + N_\bullet(t)\sigma + X^T(t)\beta}.$$

- Jako vysvětlující proměnnou $X(t)$ je možné použít např. náročnost poslední opravy.
- Pokud se hodnoty kovariáty mění jen v časech událostí, je možné snadno dosadit do logaritmické věrohodnosti a při parametrickém základním riziku maximalizovat.

Model zrychleného času

- Můžeme také předpokládat, že každá oprava či údržba a regresory způsobí, že virtuální čas plyne pomaleji nebo rychleji (Accelerated Failure Time model, AFT). Využijeme transformaci času (Lin & Ying, 1995):

$$t \to \int_0^t e^{M_\bullet(s)\rho + N_\bullet(s)\sigma + X^T(s)\beta} ds =: h(t, \boldsymbol{\beta}),$$

- kde jsme označili $\boldsymbol{\beta} = (\rho, \sigma, \beta)$. Riziková fukce pak má tvar

$$\lambda(t) = \lambda_0(h(t, \boldsymbol{\beta}))e^{M_\bullet(t)\rho + N_\bullet(t)\sigma + X^T(t)\beta}.$$

- Pokud základní riziková funkce bude konstantní, oba modely splývají.

Inference při více pozorováních

Máme-li k dispozici data o n nezávislých zařízeních, pracujeme se sdruženou věrohodností. Můžeme:

- parametrizovat základní riziko a postupovat jako výše
- nebo odhadnout základní riziko neparametricky.

Inference při více pozorováních

Máme-li k dispozici data o n nezávislých zařízeních, pracujeme se sdruženou věrohodností. Můžeme:

- parametrizovat základní riziko a postupovat jako výše
- nebo odhadnout základní riziko neparametricky.

Mějme $\lambda_i(t)$, T_{ij}, Δ_{ij}, $j = 1, ...n_i$ a $X_i(t)$ rizikovou funkci, časy událostí, indikátory oprav a hodnoty regresorů i-tého prvku.
Označíme

$$N_{ij}(t) = \Delta_{ij} I(T_{ij} \le t)$$

$$M_{ij}(t) = (1 - \Delta_{ij}) I(T_{ij} \le t)$$

$$Y_{ij}(t) = I(T_{i,j-1} < t \le T_{ij}).$$

Dostaneme log-věrohodnost

$$l = \sum_{ij} \int_0^\infty \left(\log \lambda_i(t^-) dN_{ij}(t) - Y_{ij}(t)\lambda_i(t^-) dt\right),$$

v rizikové funkci λ_i budou obsaženy počty oprav a údržeb $N_{i\bullet}$ a $M_{i\bullet}$.

Coxův model semiparametricky

- Označíme $\boldsymbol{X}_i(t) = (N_{i\bullet}(t), M_{i\bullet}(t), X_i(t))$.
- Skóre získané dosazením rizikové funkce do logaritmické věrohodnosti a derivováním podle parametrů ale závisí na $\Lambda_0(t)$.
- Tu můžeme nahradit Nelson-Aalenovým odhadem

$$\hat{\Lambda}_0(t,\beta) = \int_0^t \frac{dN_{\bullet\bullet}(s)}{\sum_{ij} e^{\boldsymbol{X}_i^T(s^-)\boldsymbol{\beta}} Y_{ij}(s)}.$$

- Po dosazení získáme skóre ve tvaru

$$U(\beta) = \sum_{ij} \int_0^\infty \left(\boldsymbol{X}_i(t^-) - \frac{\sum_{ij} \boldsymbol{X}_i(t^-) e^{\boldsymbol{X}_i^T(t^-)\boldsymbol{\beta}} Y_{ij}(t)}{\sum_{ij} e^{\boldsymbol{X}_i^T(t^-)\boldsymbol{\beta}} Y_{ij}(t)} \right) dN_{ij}(t)$$

a pro nalezení odhadů parametrů řešíme rovnice $U(\beta) = 0$.

AFT model semiparametricky

- Pro každý prvek máme transformaci času $h_i(t, \beta)$. Zavedeme transformované procesy

$$N_{ij}^*(t,\beta) = \Delta_{ij} I(h_i(T_{ij},\beta) \le t), \quad M_{ij}^*(t,\beta) = (1-\Delta_{ij}) I(h_i(T_{ij},\beta) \le t),$$

$$Y_{ij}^*(t,\beta) = I(h_i(T_{i,j-1},\beta) < t \le h_i(T_{ij},\beta)), \quad X_i^*(t,\beta) = X_i(h_i^{-1}(t,\beta)).$$

- Přesné skóre má složitější tvar, je ale možné jej nahradit přibližným (Lin & Ying, 1995) a dosadit odhad

$$\hat{\Lambda}_0(t,\beta) = \int_0^t \frac{dN_{\bullet\bullet}^*(s,\beta)}{\sum_{ij} Y_{ij}^*(t,\beta)}.$$

- Získáme

$$U(\beta) = \sum_{ij} \int_0^\infty \left(\boldsymbol{X}_i^*(t^-,\beta) - \frac{\sum_{ij} \boldsymbol{X}_i^*(t^-,\beta) Y_{ij}^*(t,\beta)}{\sum_{ij} Y_{ij}^*(t,\beta)} \right) dN_{ij}^*(t,\beta)$$

a minimalizací $\|U(\beta)\|$ opět můžeme najít odhady parametrů.

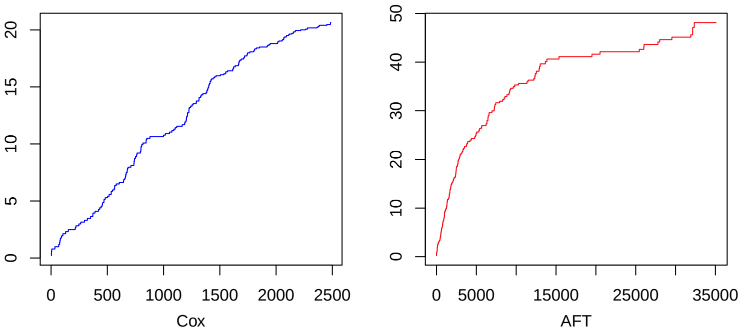
Modelování provozu čerpadla

Máme k dispozici data o provozu ropných čerpadel za několik let (Kobbacy, 1997 a Percy, 2007):

- Pro jedno čerpadlo máme podrobnější údaje o časech údržeb, oprav a o náročnosti každého zásahu v člověkohodinách. Zkusili jsme parametrické modelování životnosti při různých základních rizikových funkcích, odhadujeme ρ, σ a β v Coxově i AFT modelu.
- Pro pět čerpadel máme k dispozici jen časy oprav a údržeb. Odhadujeme ρ a σ pomocí semiparametrických metod a porovnáváme s výsledky při použití parametrizovaného základního rizika.

Výsledky viz poster.

Odhad kumulované základní rizikové funkce



Závěr

Děkuji za pozornost

Reference

[1] Kobbacy K.A.H., Fawzi B.B., Percy D.F., Ascher H.E.: *A full history proportional hazards model for preventive maintenance scheduling*, Quality and Reliability Engineering Intl. 13, 187–198, 1997.

[2] Lin D.Y., Ying Z.: *Semiparametric inference for the accelerated life model with time-dependent covariates*, Journal od Statistical Planning and Inference 44, 47-63, 1995.

[3] Percy D.F., Alkali B.M.: *Generalized proportional intensities models for repairable systems*, IMA Journal of Management Mathematics 17, 171-185, 2005.

[4] Percy D.F., Alkali B.M.: *Scheduling preventive maintenance for oil pumps using generalized proportional intensities models*, International Transactions of Operational Research 14, 547-563, 2007.