REGRESE V MODELECH OPRAV

PETR NOVÁK

novakp@karlin.mff.cuni.cz

KPMS MFF UK, ÚTIA AV ČR

Při provozu systému který podléhá opotřebení je naší snahou odhadnout jeho životnost v závislosti na dostupných informacích. Studujeme zde možnosti, jak pomocí Coxova modelu proporcionálního rizika a modelu zrychleného času popsat vliv předchozích oprav a preventivní údržby a předvádíme využití metod na datech z praxe.

Údržba a opravy

Zkoumáme data reprezentující životnost systému který podléha opotřebení. Když se systém porouchá, je nutné provést opravu, také se selhání snažíme předejít preventivními údržbami. Označíme T_j, $j = 1, ..., n$ časy oprav nebo údržeb a Δ_j indikátor zda v j-tém čase byla provedena oprava. Zavedeme čítací procesy oprav a údržeb

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

Označíme rizikovou funkci $\lambda(t) = \lim_{h\to 0} P(N(t+h)-N(t) \ge 1|\mathcal{H}(t))/h$ kde $\mathcal{H}(t)$ značí historii událostí do času t. Dále mějme kumulativní rizikovou funkci $\Lambda(t) = \int_0^t \lambda(s)ds$ a $f(t)$ a $S(t)$ příslušnou hustotu a funkci přežití. Věrohodnost lze přepsat jako

$$L = \prod_{j=1}^n \left(\frac{f(T_j^-)}{S(T_{j-1})}\right)^{\Delta_j} \left(\frac{S(T_j)}{S(T_{j-1})}\right)^{1-\Delta_j} = \prod_{j=1}^n \lambda(T_j^-)^{\Delta_j} \cdot S(T_n)$$

a log-věrohodnost má pak tvar

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

Coxův model proporcionálního rizika

Rizikovou funkci můžeme zapsat pomocí Coxova modelu. 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.

Parametrické modelování provozu čerpadla

Studujeme data o provozu čerpadla ropy (Kobbacy et al., 1997). Máme k dispozici časy oprav a preventivních údržeb a délku opravy v člověkohodinách a chceme modelovat jejich vliv na životnost. Zkusíme maximalizovat věrohodnost pro exponenciální, Weibullovo $\lambda_0(t) = a\lambda^a t^{a-1}$ a useknuté Gumbelovo $\lambda_0(t) = \lambda a^t$ základní rozdělení pro oba popsané modely. Porovnáním hodnoty věrohodnosti zjistíme, že je zde nejvyšší pro Coxův i AFT model s useknutým Gumbelovým rozdělením. Časová náročnost opravy zvyšuje riziko respektive zrychluje čas. Oprava má podle všech modelů pozitivní vliv, zajímavé je že všechny modely vyjma Gumbelova rozdělení v Coxově modelu vyhodnocují že údržba má vliv negativní.

Model	λ_0	log - věrohodnost	$e^{\hat\rho}$	$e^{\hat\sigma}$	$e^{\hat\beta}$	$\hat\lambda$	$\hat a$
	Exp.	-213.8	1.407	0.980	1.0066	0.0015	–
Cox	Weibull	-213.5	1.266	0.924	1.0064	0.0017	1.672
	Gumbel	-210.2	0.701	0.745	1.0063	0.0006	1.010
AFT	Weibull	-212.7	1.278	0.918	1.0061	0.0014	1.639
	Gumbel	-210.2	1.318	0.877	1.0050	0.0005	1.001

Inference při více pozorováních

Máme-li k dispozici data o n nezávislých zařízeních, můžeme postupovat jako výše a pracovat se součinem věrohodností, nebo můžeme odhadnout základní riziko neparametricky. Logaritmickou věrohodnost lze přepsat pomocí čítacích procesů. 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. Zavedeme

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

Pomocí $\bullet$ označíme součet přes příslušný index. Dostaneme

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

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

Neparametrický Coxův model

Pro jednoduchost 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 $\hat\Lambda_0(t, \boldsymbol{\beta}) = \int_0^t \frac{dN_{\bullet\bullet}(s)}{\sum_{ij} e^{\boldsymbol{X}_i^T(s^-)\boldsymbol{\beta}} Y_{ij}(s)}$.

Po dosazení získáme skóre ve tvaru

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

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

Neparametrický AFT model

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

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

Přesné skóre má složitější tvar, je ale možné jej nahradit přibližným (Lin & Ying, 1995) a dosadit odhad $\hat\Lambda_0(t, \boldsymbol{\beta}) = \int_0^t \frac{dN_{\bullet\bullet}^*(s,\boldsymbol{\beta})}{\sum_{ij} Y_{ij}^*(t,\boldsymbol{\beta})}$. Získáme

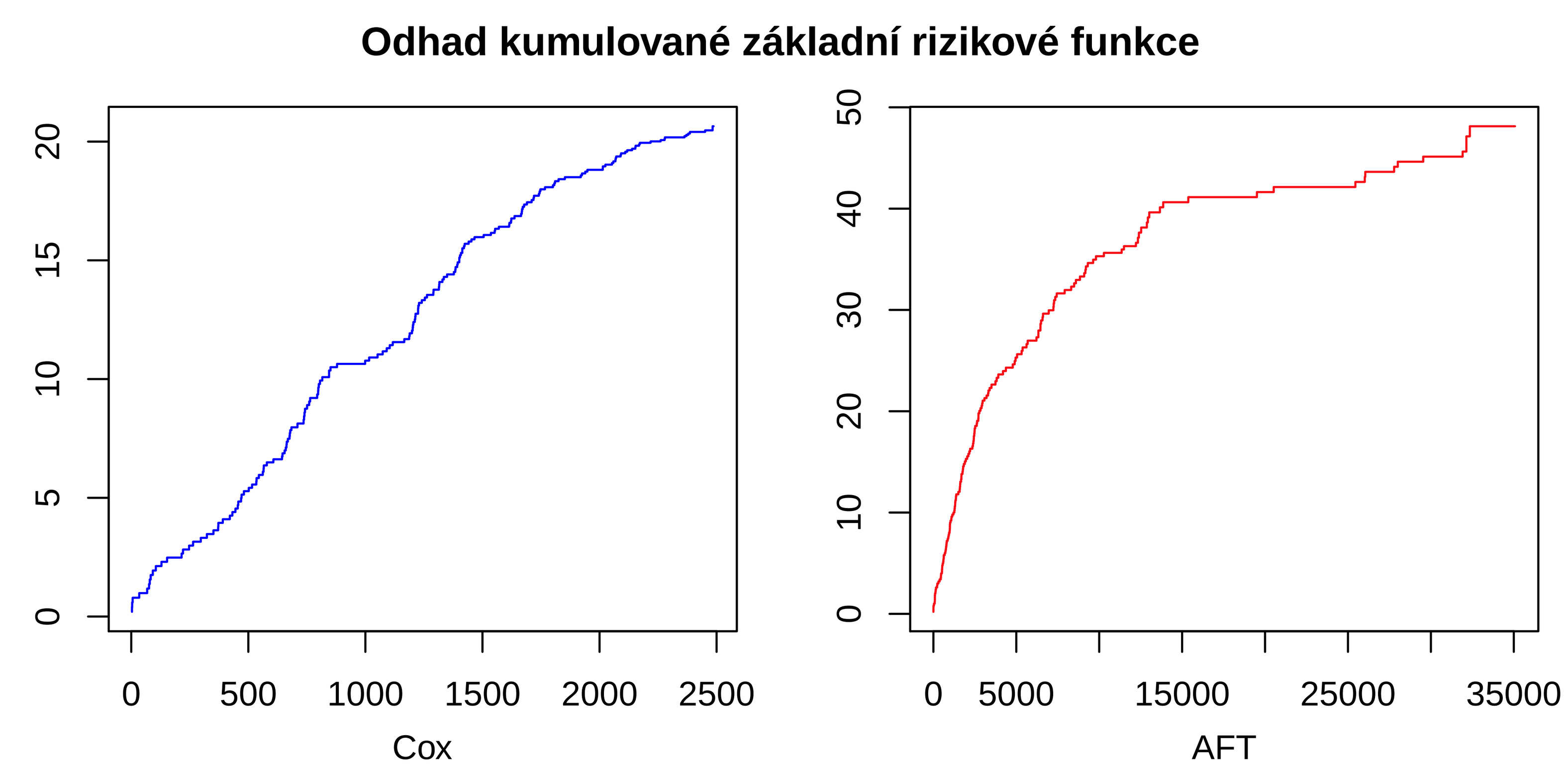
$$U(\boldsymbol{\beta}) = \sum_{ij} \int_0^\infty \left(\boldsymbol{X}_i^*(t^-, \boldsymbol{\beta}) - \frac{\sum_{ij} \boldsymbol{X}_i^*(t^-, \boldsymbol{\beta}) Y_{ij}^*(t, \boldsymbol{\beta})}{\sum_{ij} Y_{ij}^*(t, \boldsymbol{\beta})}\right) dN_{ij}^*(t, \boldsymbol{\beta})$$

a řešením $U(\boldsymbol{\beta}) = 0$ opět můžeme najít odhady parametrů.

Neparametrické modelování provozu čerpadla

Máme k dispozici časy oprav a údržeb pro pět různých čerpadel. Údaj o náročnosti opravy nebyl k dispozici u všech, takže budeme odhadovat jen regresní parametry ρ a σ. Zkusili jsme aplikovat Coxův i AFT model, jak parametricky se stejnými základními rozděleními jako výše, tak neparametricky. Vidíme, že ve všech případech oprava zvýší riziko či zrychlí plynutí času. Z parametrických modelů má nejvyšší log-věrohodnost Gumbelovo rozdělení v AFT modelu, u tohoto a neparametrických případů má údržba také negativní vliv, jinak má vliv pozitivní. Na obrázku jsou znázorněny odhady kumulované základní rizikové funkce, čas pro AFT model je v transformované škále.

Model	λ_0	log - věrohodnost	$e^{\hat\rho}$	$e^{\hat\sigma}$	$\hat\lambda$	$\hat a$
	Exp.	-880.3	0.985	1.016	0.016	–
Cox	Weibull	-880.2	0.976	1.016	0.014	1.063
	Gumbel	-880.3	0.994	1.016	0.016	0.999
AFT	Weibull	-880.2	0.980	1.015	0.014	1.038
	Gumbel	-875.1	1.022	1.036	0.013	0.999
Cox	neparam.	–	1.043	1.020	–	–
AFT	neparam.	–	1.028	1.084	–	–



Shrnutí

Zkoumali jsme metody pro modelování vlivu údržby a oprav na životnost sledovaného zařízení. V Coxově modelu působí regresory vyjadřující počet a míru zásahů a příslušné parametry multiplikativně na rizikovou funkci, v AFT modelu přímo na rychlost plynutí vnitřního času. Při parametrizaci základního rizika lze získat odhady z údajů o jednom zařízení. Pokud máme informace o více zařízeních, je možné základní riziko odhadnout neparametricky.

Poděkování

Tato práce byla podporována granty SVV 261315/2012 a MŠMT ČR 1M06047.

Reference

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

[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] 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.