

Model ryzyka oparty na regresji rangowej¹

Analiza historii zdarzeń pozwala konstruować modele prognostyczne i przewidywać rozwój badanych zjawisk. Modele prognostyczne wykorzystywane są m.in. do oceny ryzyka pogorszenia się zdrowia lub śmierci poszczególnych pacjentów w wyniku operacji serca. Modele ryzyka są budowane na podstawie danych analizy przeżyć (ang. *survival analysis*) [1]. W tym celu wykorzystuje się powszechnie model Cox'a [2]. W prezentowanym wystąpieniu przeanalizowane zostaną możliwości wykorzystania regresji rangowej do modelowania ryzyka [3].

Dane analizy przeżyć przedstawiane są tradycyjnie jako zbiór C o poniższej strukturze [1]:

$$C = \{(\mathbf{x}_j, t_j, \delta_j)\} \quad (j = 1, 2, \dots, m) \quad (1)$$

gdzie $\mathbf{x}_j = [x_{j1}, \dots, x_{jn}]^T$ jest tzw. wektorem cech x_i opisującym stan j -tego pacjenta O_j , t_j jest czasem obserwacji tego pacjenta, δ_j jest wskaźnikiem cenzorowania ($\delta_j \in \{0, 1\}$). Wartość $\delta_j = 0$ wskazuje na koniec obserwacji danego pacjenta przed zaistnieniem zdarzenia, natomiast wartość $\delta_j = 1$ wskazuje na koniec obserwacji związany z badanym zdarzeniem (np. śmiercią).

W rangowych modelach ryzyka wykorzystujemy relację porządkową $R_j \prec R_k$ oznaczającą, że ryzyko R związane z pacjentem O_j jest mniejsze niż to związane z pacjentem O_k :

$$(R_j \prec R_k) \iff (\delta_k = 1 \text{ i } t_j > t_k) \quad (2)$$

Projektując rangowy model ryzyka poszukujemy takiego wektora parametrów \mathbf{w} , dla którego odwzorowanie liniowe $y(\mathbf{w}) = \mathbf{w}^T \mathbf{x}$ możliwie dobrze zachowuje poniższą implikację:

$$(R_j \prec R_k) \Rightarrow (\mathbf{w}^T \mathbf{x}_j < \mathbf{w}^T \mathbf{x}_k) \quad (3)$$

Zagadnienie projektowania rangowego modelu ryzyka można sprowadzić do zagadnienia liniowej separowalności dwu zbiorów oraz zastosować wypukłą i od-cinkowo-liniową (CPL) funkcję kryterialną do wyznaczenia optymalnego wektora parametrów \mathbf{w} .

Literatura

- [1] J. P. Klein, M. L. Moeschberger, *Survival Analysis*, Techniques for Censored and Truncated Data, Springer, NY 1997.
- [2] O. D. Allison, *Survival Analysis Using the SAS System. A Practical Guide*, SAS Institute Inc., Cary, NC, USA 1995.
- [3] L. Bobrowski, *Ranked modeling of risk on the basis of survival data*, ICSMRA2007, Lisbon.

¹ Praca statutowa 16/St/2007 IBIB PAN.