

## Modele prognostyczne analizy przeżyć oparte na regresji rangowej

Nowe leki jak też nowe sposoby leczenia oceniane są powszechnie w oparciu o metody analizy przeżyć (ang. *survival analysis*) [1]. Statystyczne metody analizy przeżyć służą głównie badaniu i modelowaniu historii zdarzeń chorobowych [2]. W oparciu o te metody można budować modele prognostyczne opisujące rozwój choroby u wybranych pacjentów. Modele takie powinny np. poprawnie prognozować czas życia pacjenta po operacji serca.

Parametry modeli analizy przeżyć estymowane są na podstawie zbiorów danych  $C$  o charakterystycznej strukturze z tzw. cenzorowaniem (ang. *censoring*) [3]

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

gdzie  $\mathbf{x}_j = [x_{j1}, \dots, x_{jn}]^T$  jest tzw. wektorem cech (atrybutów)  $x_i$  opisującym stan  $j$ -tego pacjenta  $O_j$ ,  $t_j$  jest czasem obserwacji tego pacjenta,  $\delta_j$  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 (śmiercią). W rangowych modelach prognostycznych wykorzystujemy poniższą relację [4]

$$\mathbf{x}_j < \mathbf{x}_k \iff \mathbf{x}_k \text{ ma dłuższy czas życia niż } \mathbf{x}_j \quad (2)$$

**Definicja 1.** Wektor  $\mathbf{x}_k$  (pacjent  $O_k$ ) ma dłuższy czas życia niż  $\mathbf{x}_j$  ( $\mathbf{x}_j < \mathbf{x}_k$ ), gdy spełniona jest relacja

$$\delta_j = 1 \text{ oraz } t_j < t_k. \quad (3)$$

Wykorzystując dane  $C$  (1) oraz relację (2) można skonstruować liniowe odwzorowanie rangowe, które zachowuje w możliwie dużym stopniu relację (2) na prostej  $y(\mathbf{w})$  (4) [5].

$$y_j = y_j(\mathbf{w}) = \mathbf{w}^T \mathbf{x}_j, \quad (4)$$

gdzie  $\mathbf{w} = [w_1, \dots, w_n]^T$  jest wektorem parametrów.

Odwzorowanie rangowe może opisywać *kolejność umierania*. W oparciu o to odwzorowanie można prognozować czas przeżycia poszczególnych pacjentów, jak również wyznaczyć takie cechy  $x_i$ , które są najbardziej znaczące w prognozie.

### Literatura

- [1] J. P. Klein, M. L. Moeschberger, *Survival Analysis*, Springer, New York 1997.
- [2] E. Frątczak, U. Gach-Ciepiela, H. Babiker, *Analiza historii zdarzeń — elementy teorii, wybrane przykłady zastosowań*, Wydawnictwa AGH, Warszawa 2005.
- [3] O. D. Allison, *Survival Analysis Using the SAS System*, Cary, 1995.
- [4] L. Bobrowski, T. Łukaszuk, *Ranked linear modeling in survival analysis*, w: Lecture notes of the ICB Seminar „Statistics and Clinical Practice”, Warsaw, June 2005, 61–67.
- [5] L. Bobrowski: *Eksploracja danych oparta na wypukłych i odcinkowo-liniowych funkcjach kryterialnych*, Wydawnictwa Politechniki Białostockiej, 2005.