

Ichino's feature selection method for symbolic data

The Ichino's (1994) method serves to select the effective (non-noisy) features from primary set of symbolic variables described a set of objects. The method is appropriate for feature selection for objects set described by symbolic variables of the types: interval-valued, multinomial and multinomial with weights variables. The method is basing on the *Cartesian space model* – CSM (Ichino, Yaguchi [1994]) and the *Pretended simplicity theorem* based on the mutual neighborhood graphs – MNG (Ichino [1986], Ichino, Sklansky [1985]) defined on the CSM. Due to an exponential calculation complexity the Ichino's method is recommended for relatively small variables set.

The algorithm of Ichino's feature selection method is as follows (Ichino [1994]):

Step 1. Symbolic data table containing n objects described by m symbolic variables.

Step 2. Find all mutual neighbor graphs for each combination of candidate-variables.

Step 3. Calculate the number of mutual neighbor pairs inside graphs that is $\binom{n}{2} = \frac{n \cdot (n-1)}{2}$.

Step 4. Find the combination of candidate-variables for which the sum calculated in step 2 is the highest.

Step 5. Choose the variable combination for which the increment of value calculated in step 3 – in relation to $k-1$ candidate-variables – is the highest.

The defined set of effective variables (the combination most differentiating the set of objects) can be included in further analysis.

References

- Ichino, M. (1994), *Feature selection for symbolic data classification*, In: E. Diday, Y. Lechevallier, P.B. Schader, B. Burtschy (Eds.), *New Approaches in Classification and data analysis*, Springer-Verlag, pp. 423-429.
- Bock, H.H., Diday, E. (Eds.) (2000), *Analysis of symbolic data. Explanatory methods for extracting statistical information from complex data*, Springer-Verlag, Berlin.
- Diday, E., Noirhomme-Fraiture, M. (Eds.) (2008), *Symbolic Data Analysis with SODAS Software*, John Wiley & Sons, Chichester.
- Ichino, M. (1986), *Pattern classification based on the Cartesian join system: A general tool for feature selection*, IEEE Int. Conf. on SMC, Atlanta.
- Ichino, M., Yaguchi, H. (1994), *Generalized Minkowski Metrics for Mixed Feature-Type Data Analysis*, "IEEE Transactions on Systems, Man, and Cybernetics", Vol. 24, No. 4, pp. 698-707.
- Ichino, M., Sklansky, J. (1985), *The relative neighborhood graph for mixed feature variables*, Pattern Recognition, Vol. 18, No. 2, pp. 161-167.