

Validation of soft classifiers for cells and tissues

C. Beleites^{1,2}, C. Krafft², J. Popp^{2,3}, and V. Sergo¹

¹CENMAT and Dept. of Industrial and Information Engineering, Univ. Trieste, Trieste/Italy

²Institute of Photonic Technology, Jena/Germany

³Institute of Physical Chemistry and Abbe Center of Photonics, University Jena/Germany

Medical diagnosis of cells and tissues is an important aim in biospectroscopy. The data analytical task involved frequently is classification. Classification traditionally assumes both reference and prediction to be *hard*, i.e. stating exactly one of the defined classes. Like fuzzy cluster analysis, soft classification uses partial memberships rather than hard labels, thus expressing uncertainty or mixed cell populations.

Many classification methods produce soft output, e.g. posterior probabilities. Some methods also use partial training labels. Yet, for medical diagnostic applications it is even more important to include soft samples into the model validation. Excluding ambiguous samples means retaining only clear (i.e. easy) cases. Such a test set is not representative of the original unfiltered population, and creates a risk of obtaining overly optimistic estimates of the model performance. While mixed populations and uncertain reference are more common for tissue classification, a second problem concerns cells as well: The classical calculation of performance measures such as sensitivity, specificity, and predictive values often comprises a lossy translation of e.g. posterior probabilities into hard class labels, a practice long criticized in the statistical community (e.g. [1]) which also constitutes a major hindrance for the optimization of classifiers.

We introduce a framework to calculate classifier performance measures for samples with soft reference and prediction which also avoids the dichotomization loss. Briefly, if the partial memberships are interpreted as uncertainty, best and worst case as well as expected performance are obtained via the weak, strong and product conjunction [2]. For the mixture interpretation, weighted versions of well-known regression performance measures like mean absolute and root mean squared errors result. A ready-to-use implementation (R package) is available at <http://softclassval.r-forge.r-project.org>.

As real world example, we classify 37 015 Raman (thereof 55% soft) spectra of 80 brain tumor patients into “normal”, “low grade”, and “high grade” tissue morphologies in order to delineate excision borders during surgical treatment of the tumors. Thus, borderline cases are our actual target samples.

Financial support by the Associazione per i Bambini Chirurgici del Burlo (IRCCS Burlo Garofolo Trieste) and of the European Union via the Europäischer Fonds für Regionale Entwicklung (EFRE) and the “Thüringer Ministerium für Bildung, Wissenschaft und Kultur” (Project: B714-07037) is highly acknowledged.

References

- [1] J. Cohen, The cost of dichotomization, *Applied Psychological Measurement*, 7, 249 – 253 (1983).
- [2] S. Gottwald, Many-valued logic. In E. N. Zalta (ed.): "The Stanford Encyclopedia of Philosophy" (Spring 2010 ed.).