The Relation between DSEA and IBU

Mirko Bunse

Supplementary Material 2018

When DSEA (the Dortmund spectrum estimation algorithm) employs a Naive Bayes classifier trained on a single nominal attribute, it is equivalent to IBU (the iterative Bayesian unfolding). In the sense that DSEA is open to other classification algorithms and arbitrary mixtures of multiple nominal and numerical attributes, it is thus more general than IBU.

To see this relation between the two algorithms, let us recap the reconstruction rule of DSEA presented in our paper [1] as equation (21).

$$\hat{\mathbf{f}}_i^{\textsc{Dsea}} = \frac{1}{N} \sum_{n=1}^{N} c_{\mathcal{M}}(\, i \,|\, \mathbf{x}_n)$$

Moreover, consider the confidence values defined for a Naive Bayes classifier, which applies Bayes' theorem to model the posterior probabilities of labels.

$$\begin{aligned} c_{\mathcal{M}}(\, i \,|\, \mathbf{x}_n) &= \hat{\mathbb{P}}\,(\, Y \equiv i \mid X = \mathbf{x}_n) \\ &= \frac{\hat{\mathbb{P}}\,(\, X = \mathbf{x}_n \mid Y \equiv i\,) \cdot \hat{\mathbb{P}}\,(\, Y \equiv i\,)}{\sum_{i'=1}^{I} \hat{\mathbb{P}}\,(\, X = \mathbf{x}_n \mid Y \equiv i'\,) \cdot \hat{\mathbb{P}}\,(\, Y \equiv i'\,)} \end{aligned}$$

Now let X consist of a single nominal attribute with J discrete values. Without loss of generality, this means $\mathcal{X} = \{1, 2, \ldots, J\}$. Moreover, let $\mathcal{D}_j = \{\, \mathbf{x} \in \mathcal{D}_{\mathrm{obs}} \mid \mathbf{x} = j\,\}$ be the set of observed examples with the attribute value j. Grouping the N observations $\mathbf{x}_n \in \mathcal{D}_{\mathrm{obs}}$ by their attribute values corresponds to a mere re-ordering of the observations.

$$\begin{aligned} \hat{\mathbf{f}}_i^{\textsc{Dsea}} &= \frac{1}{N} \sum_{n=1}^{N} \frac{\hat{\mathbb{P}}\,(\, X = \mathbf{x}_n \mid Y \equiv i\,) \cdot \hat{\mathbb{P}}\,(\, Y \equiv i\,)}{\sum_{i'=1}^{I} \hat{\mathbb{P}}\,(\, X = \mathbf{x}_n \mid Y \equiv i'\,) \cdot \hat{\mathbb{P}}\,(\, Y \equiv i'\,)} \\ &= \frac{1}{N} \sum_{j=1}^{J} \sum_{\mathbf{x} \in \mathcal{D}_j} \frac{\hat{\mathbb{P}}\,(\, X = \mathbf{x} \mid Y \equiv i\,) \cdot \hat{\mathbb{P}}\,(\, Y \equiv i\,)}{\sum_{i'=1}^{I} \hat{\mathbb{P}}\,(\, X = \mathbf{x} \mid Y \equiv i'\,) \cdot \hat{\mathbb{P}}\,(\, Y \equiv i'\,)} \end{aligned}$$

In this form, all terms in the inner sum are equal to each other because all $\mathbf{x} \in \mathcal{D}_j$ have the same value. Since this value is j, $X = \mathbf{x}$ can be written as $X \equiv j$. Moreover, let N_j be the number of examples in $\mathcal{D}_j$. The nested sum can then be simplified.

$$\hat{\mathbf{f}}_i^{\text{DSEA}} \quad = \quad \frac{1}{N} \sum_{j=1}^{J} N_j \cdot \frac{\hat{\mathbb{P}}\,(\, X \equiv j \,|\, Y \equiv i \,) \cdot \hat{\mathbb{P}}\,(\, Y \equiv i \,)}{\sum_{i'=1}^{I} \hat{\mathbb{P}}\,(\, X \equiv j \,|\, Y \equiv i' \,) \cdot \hat{\mathbb{P}}\,(\, Y \equiv i' \,)}$$

Recall that IBU estimates $\mathbf{g}_j$ from the relative number $\frac{N_j}{N}$. Moreover, each conditional probability $\hat{\mathbb{P}}\,(\, X \equiv j \,|\, Y \equiv i \,)$ is given by the corresponding component $\mathbf{R}_{ij}$ of the detector response matrix. $\hat{\mathbb{P}}\,(\, Y \equiv i \,)$ is a component of the prior on $\mathbf{f}$, which is equivalently written as $\hat{\mathbf{f}}_i^{(0)}$. Rewriting the DSEA update rule under these considerations yields the update rule of IBU from Equation (18). DSEA is thus equivalent to IBU, if a single nominal feature is used to train a Naive Bayes classifier.

$$\hat{\mathbf{f}}_i^{\text{DSEA}} \quad = \quad \sum_{j=1}^{J} \frac{R_{ij}\, \hat{\mathbf{f}}_i^{(0)}}{\sum_{i'=1}^{I} \; R_{i'j}\, \hat{\mathbf{f}}_{i'}^{(0)}} \cdot \mathbf{g}_j \quad = \quad \hat{\mathbf{f}}_i^{\text{IBU}}$$

This equivalence does not hold if multiple features or a single numerical feature are used in DSEA because the probabilities are computed differently, then. Moreover, the algorithms are not equivalent if another classifier than Naive Bayes is employed. In the sense that DSEA is able to use multiple features to train an arbitrary classifier, DSEA is thus more general than the Iterative Bayesian Unfolding. This relation is first given here, not being published before.

References

[1] Mirko Bunse, Nico Piatkowski, Tim Ruhe, Wolfgang Rhode, and Katharina Morik. Unification of deconvolution algorithms for cherenkov astronomy. In *5th IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 2018.