1145-00-2757 **Edgar E. Robles** (edgar.roblesarias@ucr.ac.cr), **Ching Pui Wan**, **Fatima Zaidouni*** (fzaidoun@u.rochester.edu), **Joanne Beckford** and **Aliki Mavromoustaki**. *Threshold optimization in multiple binary classifiers for extreme rare events using predicted positive data.*

Classification on imbalanced datasets is a challenging problem where a high rate of correct detection is required in the minority class. We analyze the output of binary classification models used by Google, where the inputs are documents categorized as either predicted positive or negative against a certain threshold. In rare-event problems, positives have a prevalence of around 0.1% and it is expensive to estimate all documents. Therefore, the problem is reformulated using the correct labels [true positive (TP) or false positive (FP)] on a sample of the predicted positives, as determined by human raters. It is important to pick an operating point (OP) on the TP/FP fitted curve whose position is adjusted to return the cost for one additional TP document in terms of the number of FP. We propose two solutions to select an optimal OP by maximizing the area under the curve (AUC): a graph-based and an analytic approach. The graph-based approach constructs a graph to select an optimal path in the threshold space that is then converted to a curve in the TP/FP space. The analytic approach estimates the AUC by minimizing a cost function. Our approaches improve over existing solutions by relating the TP/FP space to the threshold space and offer a business interpretation to the OP. (Received September 25, 2018)