CMSD 636: Annotated Bibliography (Individual Submission)
Student Name: Sourajit Saha Student ID: JI60692
Group (#2): SeeBeL Spring 2023

Cost Curves [1] is an essential visualization method to uncover machine learning model performance over a range of class distributions. The idiom for cost curve is line plot and the authors plot [1] probability cost of a wide range of sampling distribution over a dataset and their corresponding normalized expected cost. Moreover, to encode data, the authors have used horizontal and vertical spatial positions and point marks for the user to find trend and correlation. However, generalizing the cost curve [1] method to semantic segmentation model is computationally (both memory and time) expensive since (1) cityscapes has 19 trainable classes and the cost curve method uses only two classes to sample a wide range of distribution samples and (2) the complexity of a segmentation model is much higher than that of a classification model. Furthermore, from a visualization perspective the cost curve [1] (shown in Fig. 1) is not easily interpretable and it takes domain specific knowledge on the user's end and added cognitive load to compare a correlation between class distribution and model performance.. Therefore, we only propose to report model performance and class distributions of the dataset using Bubble Charts to compare correlation between them in our project.
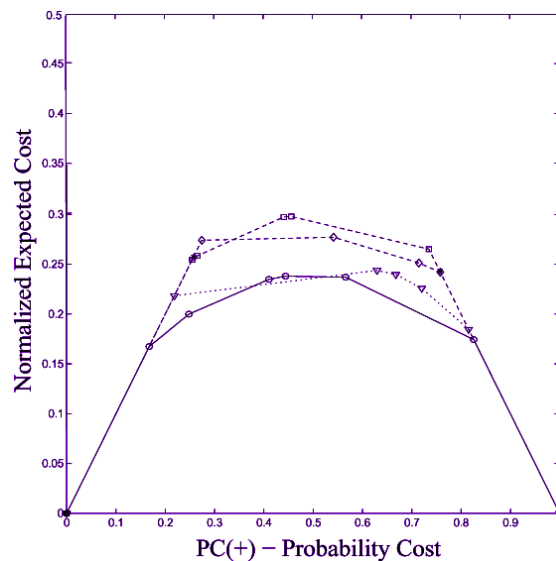

Figure 1: Cost Curve [1] depicting how class distribution and performance is correlated.

References:

1. C. Drummond and R. C. Holte, "Cost curves: An improved method for visualizing classifier performance," Machine learning, vol. 65, pp. 95–130, 2006.