Supplementary Materials: Agglomerative Clustering with Threshold Optimization via Extreme Value Theory

Chunchun Li [1], Manuel Günther [2], Akshay Raj Dhamija [1,†], Steve Cruz [1,†], Mohsen Jafarzadeh [1,†], Touqeer Ahmad [1,†] and Terrance E. Boult [1,*]

1. Supplementary Material

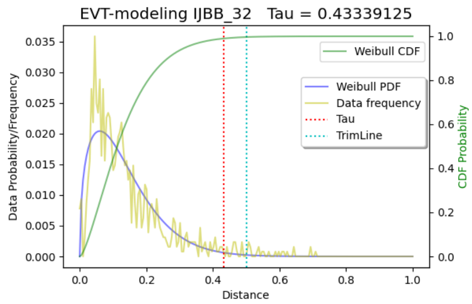
1.1. Theory

In this Supplementary Material, we provide additional details on theory and on experimental results by including timing, different metrics, and other comparison algorithms.

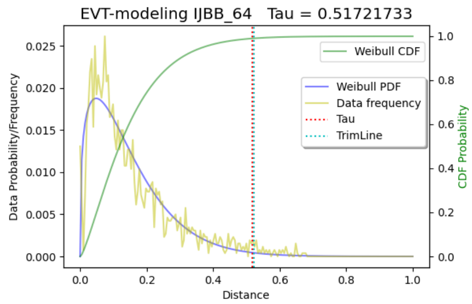
We start by recalling the Fisher–Tippet theorem, also known as the statistical Extreme Value Theory (EVT), which we use in the proof of our main theorem in the main paper. Just as the Central Limit Theorem dictates that the random variables generated from certain stochastic processes follow Gaussian distributions, EVT dictates that given a well-behaved initial distribution of values, e.g., a distribution that is continuous and has an inverse, the distribution of the maximum/minimum values can assume only limited forms.

Theorem 0 (Fisher–Tippet Theorem [1]:)**.** *Let* $(v_1, v_2, \ldots)$ *be a sequence of i.i.d samples. Let* $\zeta_n = \max\{v_1, \ldots, v_n\}$*. If a sequence of pairs of real numbers* (a_n, b_n) *exists, such that each* $a_n > 0$ *and* $\lim_{z\to\infty} P\big(\frac{\zeta_n - b_n}{a_n} \leq z\big) = F(z)$ *then if* F *is a non-degenerate distribution function, it belongs to the Gumbel, the Fréchet, or the Reversed-Weibull family.*

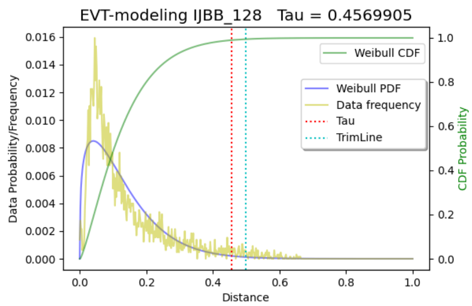
This Extreme-value theorem is widely used in many fields [1], such as manufacturing, e.g., estimating time to failure, natural sciences, e.g., estimating 100 or 500-year flood levels, and finance, e.g., portfolio risks. EVT has recently been (re)introduced and applied in recognition, machine learning, and computer vision [2–4].



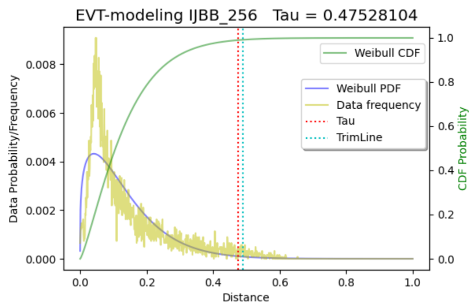
(a) IBJJ-32



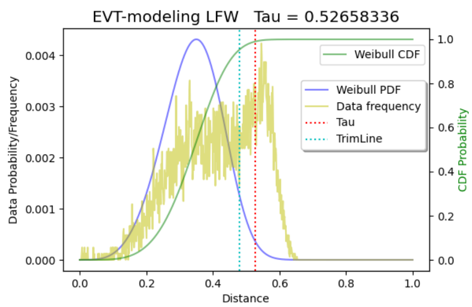
(b) IBJJ-64



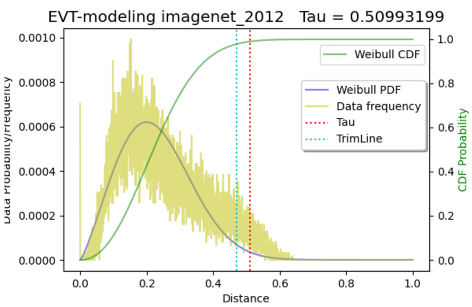
(c) IBJJ-128



(d) IBJJ-256



(e) LFW-Trimmed



(f) Imagenet-2012

Figure S1. WEIBULL FIT AND RESULTING τ. Example plots showing Weibull fit and resulting threshold τ estimated from the distribution of ANN distances. In this figure, we show a histogram of raw data from the named dataset, the resulting Weibull fit, its CDF, and the resulting τ_w for 98% of the data and 99% confidence, except for ImageNet and LFW where we show the "robust" version based on the mode heuristic which results in using only 88% and 63% of the data, respectively. For LFW, the resulting fit is quite different from the example fit for all LFW data, which is shown in the main paper. These plots show the trim-line of what data were ignored in fitting the Weibull.

For those wondering about the impact of mixing linkages from different clusters, the underlying i.i.d, nature can be sampling from a mixture and does not assume uniformity. The free sequences b_i a_i in the Fisher–Tippet theorem can allow different subsequences and can be viewed as normalizing for each of the different underlying classes if some classes have a larger average distance between points. However, it is worth noting that as an asymptotic theorem, increasing mixture components could easily delay convergence, so there needs to be enough samples from each mixture element for the theorem to apply. Thus singletons, which can be viewed as undersampled clusters, limit the convergence to the underlying Weibull.

Some example fits were in the main paper with more examples shown in Figure S1. As you can see, most of the fits are quite good. We note that all of the IJB-B data have some singletons—clusters of only one point, which our theory views as outliers. For example, IJB-B-512 has about 10% singleton classes, i.e., outliers, but we do not presume such knowledge, and so the base algorithm only trims a small amount of data. In all but one fit, we trim only 2%. For LFW, the number of singletons is more than 50% of the "clusters"—our mode-based heuristic detects outliers. The difficulty of estimating the actual fraction of outliers is part of why we developed the scaling heuristics in the main paper.

References

1. Kotz, S.; Nadarajah, S. *Extreme Value Distributions: Theory and Applications*; World Science: Singapore, 2001.
2. Scheirer, W.J.; Rocha, A.; Micheals, R.J.; Boult, T.E. Meta-Recognition: The Theory and Practice of Recognition Score Analysis. *TRansactions Pattern Recognit. Mach. Intell.* **2011**, *33*, 1689–1695.
3. Scheirer, W.; Jain, L.; Boult, T. Probability Models for Open Set Recognition. *Trans. Pattern Recognit. Mach. Intell.* **2014**, *36*, 2317–2324.
4. Carpentier, A.; Valko, M. Extreme bandits. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 8–13 December 2014. Available online: https://papers.nips.cc/paper/2014/file/8c7bbbba95c1025975e548cee86dfadc-Paper.pdf (accessed on 10 April 2022).