

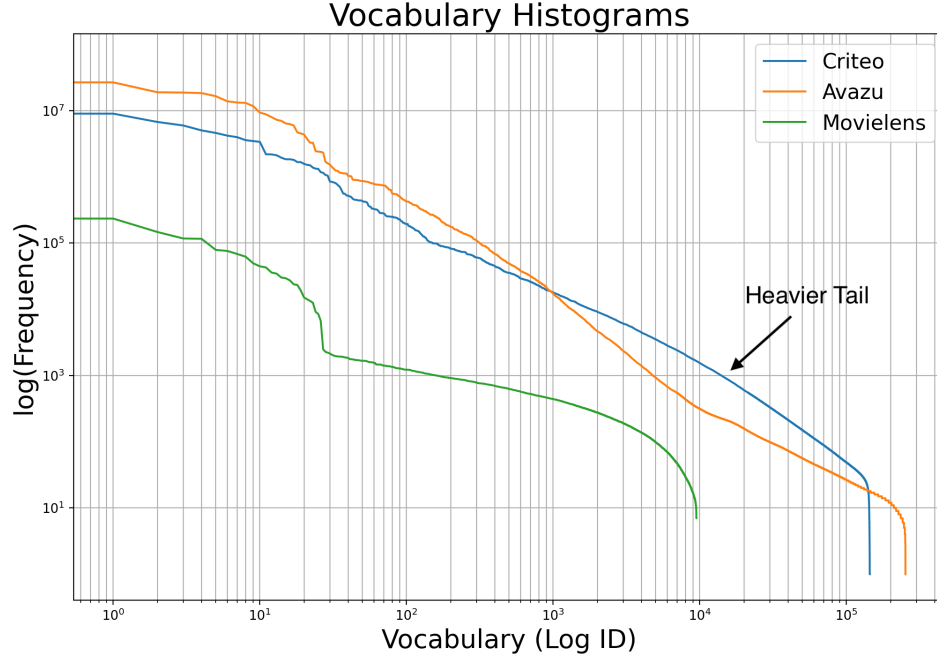


Figure 1: Vocabulary distribution for all categorical feature values (merged) in Criteo, Avazu, and Movielens. All of these datasets display power law behavior, but Criteo has a much heavier tail.

Table 1: AUC of embedding methods on click-through data. We give results for several points on the parameter-accuracy tradeoff, listed as sub-headings under the dataset that describe the total embedding size. Our proposed methods (listed below the bar) outperform the strongest baseline at the ** 0.01 and * 0.05 significance level, Welch’s t-test.

Method	Criteo			Avazu			Movielens-1M		
	25MB	12.5MB	2.5MB	32.4MB	3.24MB	324kB	1.6MB	791kB	158kB
Collisionless	80.70	–	–	77.35	–	–	88.72	–	–
Hashing Trick	80.21	79.98	79.44	77.24	76.71	75.10	85.37	83.00	77.07
Hash Embedding	79.82	80.34	80.40	77.29	76.89	76.86	88.28	87.71	86.05
HashedNet	80.54	80.50	80.42	77.30	77.19	76.89	87.92	87.92	87.31
ROBE-Z Embedding	80.55	80.52	80.41	77.32	77.18	76.88	88.05	87.92	87.40
PQ Embedding	80.29	80.25	79.82	77.39	77.11	76.48	88.49	88.38	87.32
QR Embedding	80.29	80.21	79.74	76.98	76.61	75.86	88.49	88.63	88.22
Multiplex Hash Trick	80.43	80.47	80.49**	77.35	77.18	76.86	87.74	86.93	82.00
Multiplex HashEmb	80.52	80.42	80.31	77.28	77.16	76.82	88.79**	88.74*	87.66
Multiplex HashedNet	80.53	80.55	80.48**	77.30	77.23*	76.94**	87.95	87.93	87.57
Multiplex ROBE-Z	80.57	80.56	80.50*	77.32	77.24**	76.94**	88.00	88.03	87.62
Multiplex PQ	80.43	80.49	80.54**	77.39	77.29**	76.96**	88.81**	88.69*	88.09
Multiplex QR	80.22	80.28	80.48**	77.29	77.16	76.90	88.83**	88.53	88.27**

1 **Significance tests for results in Table 1:** When we compared the multiplexed methods against the
2 corresponding non-multiplexed baseline (rather than the strongest baseline), the following differences
3 were significant with $p < 0.05$ (Welch’s t-test).

4 **Criteo:** All results in the 2.5 MB column. Multiplex Hashing Trick and Multiplex PQ in the 12.5
5 MB column. Multiplex Hashing Trick, Multiplex HashEmb, Multiplex PQ, and Multiplex QR in the
6 25MB column.

7 **Avazu:** All results in the 324 kB and 3.24 MB columns. Multiplex QR in the 32.4 MB column.

8 **Movielens:** All results in the 158 kB column. All results in the 791 kB column, except for Multiplex
9 ROBE-Z. Multiplex HashEmb, Multiplex PQ, and Multiplex QR in the 1.6 MB column.

10