Supplementary Material

Yusheng Xie Zhengzhang Chen Kunpeng Zhang Md. Mostofa Ali Patwary
Yu Cheng Haotian Liu Ankit Agrawal Alok Choudhary

This document provides the evidence against a non-uniform Day-Of-Week (DOW) / Hour-Of-Day (HOD) pattern existing in user feedback distributions, given MBT efforts, which is a claim made in section 4.1 from our paper manuscript. Goodness-of-fit tests are performed to support our claim.

1 Bucketing MBT efforts and feedback.

The bucketing process for DOW feedback distribution takes a stream of feedback data over the same 60-day period used in Figure 3 and Figure 4 of our paper manuscript, and it creates a histogram of 7 bins, each of which corresponds to a day in the week. The value at each bin of the histogram is the quantity of "feedback per effort". Using "feedback per effort", instead of the total amount of feedback in that bin, can isolate the effect of effort that might act upon feedback DOW pattern. A similar process is carried out for HOD pattern, where 24, instead of 7, bins are created. We uniformly sample 1,000 efforts across the four categories in our dataset. But for each effort in the sample, all its feedback is taken into consideration.

2 χ^2 tests for patterns in feedback distribution.

A series of Pearson's χ^2 tests are performed on values from the DOW / HOD histogram to test whether the DOW / HOD pattern is statistically significant. Let p_i / q_i denote the true value of percentage at the ith bin of the DOW / HOD histogram. Then consider the following sets of pairs of null hypotheses and alternative hypotheses.

$$\mathcal{H}_{DOW} = \left\{(H_0^i : p_i = p_j, H_1^i : p_i \neq p_j) | i, j \in \{1, ..., 7\}, i < j\right\}$$

$$\mathcal{H}_{HOD} = \left\{(H_0^i : q_i = q_j, H_1^i : q_i \neq q_j) | i, j \in \{0, 1, ..., 23\}, i < j\right\}$$

It is not hard to work out the number of pairs of hypotheses in either $\mathcal{H}_{DOW}$ or $\mathcal{H}_{HOD}$.

$$|\mathcal{H}_{DOW}| = \frac{7!}{2!(7-2)!} = 21$$

$$|\mathcal{H}_{HOD}| = \frac{24!}{2!(24-2)!} = 276$$

Table 1: Pearson's χ^2 tests for DOW pattern.

Hypotheses	**χ^2 value**	**d.o.f.**	**χ^2 at 95%**
H_0^1 vs H_1^2	1.324	1	3.84
H_0^1 vs H_1^3	3.667		
H_0^1 vs H_1^4	3.452		
H_0^1 vs H_1^5	**6.533**		
H_0^1 vs H_1^6	**5.657**		
H_0^1 vs H_1^7	**6.766**		
H_0^2 vs H_1^3	0.6154		
H_0^2 vs H_1^4	1.520		
H_0^2 vs H_1^5	0.4737		
H_0^2 vs H_1^6	0.2444		
H_0^2 vs H_1^7	0.5447		
H_0^3 vs H_1^4	0.05882		
H_0^3 vs H_1^5	0.002247		
H_0^3 vs H_1^6	0.5447		
H_0^3 vs H_1^7	0.05882		
H_0^4 vs H_1^5	0.002247		
H_0^4 vs H_1^6	0.08571		
H_0^4 vs H_1^7	0.03810		
H_0^5 vs H_1^6	0.002532		
H_0^5 vs H_1^7	0.08978		
H_0^6 vs H_1^7	0.6910		

We compute the Pearson's χ^2 test statistic (i.e. the χ^2 value) for all hypotheses in $\mathcal{H}_{DOW}$ and $\mathcal{H}_{HOD}$. Table 1 contains the summary of our findings in $\mathcal{H}_{DOW}$. The only 3, out of 21, bold-faced entries in Table 1 signify the cases where the null hypothesis (i.e. the feedback activity during the two days are the same) is rejected with P-value less than 0.05 (i.e. 95% confidence). Similar excises are carried out for analyzing HOD pattern, where only 18 out of 276 hypothesis tests reject the null hypothesis. Such results provide us strong evidence for the independence of DOW/HOD's influence over user feedback activities once given MBT efforts.