Causal Analysis of the Unsatisfying Experience in Realtime Mobile Multiplayer Games in the Wild

Yuan Meng, Shenglin Zhang,Zijie Ye,
Benliang Wang, Zhi Wang, Yongqian Sun, Qitong Liu, Shuai Yang, Dan Pei

Tsinghua University

Tencent Games

Background

Unsatisfying user experience in the Realtime Mobile Multiplayer Games(RMMG) is frustrating, which could lead to the loss of customers for the game company. We focus on two metrics in the paper:

- **LRC**: the number of Location Resynchronizations experienced by a player in a session
- **AQ**: whether a player quits the session abnormally

Fig. 1: Location Resynchronizations and Abnormal Quit

Unsatisfying user experience in the whild

Based on 12 million real game sessions from a top-tier RMMG, We observe that 13% of the game sessions suffer from at least one location resynchronization, and 7.12% have been aborted abnormally before the end of the game.

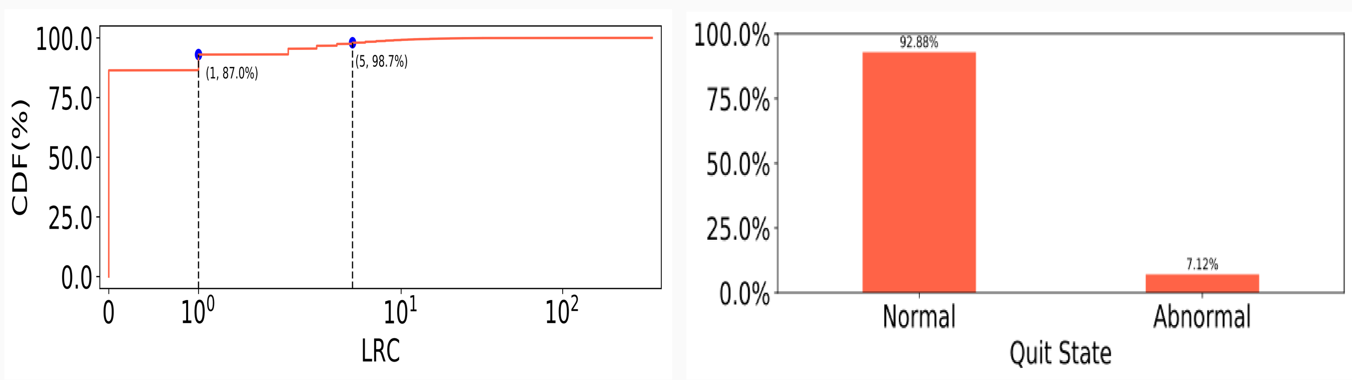


Fig. 2: User experience in the whild

ExCause: a general causal analysis framework of user experience

To study the causes of the unsatisfying experience, our paper proposes, *ExCause*, a general causal analysis framework to systematically analyze historical game session records to 1) obtain context factors that cause unsatisfying RMMG experience and 2) to recommend adjustments with quantified QoE improvement expectation.

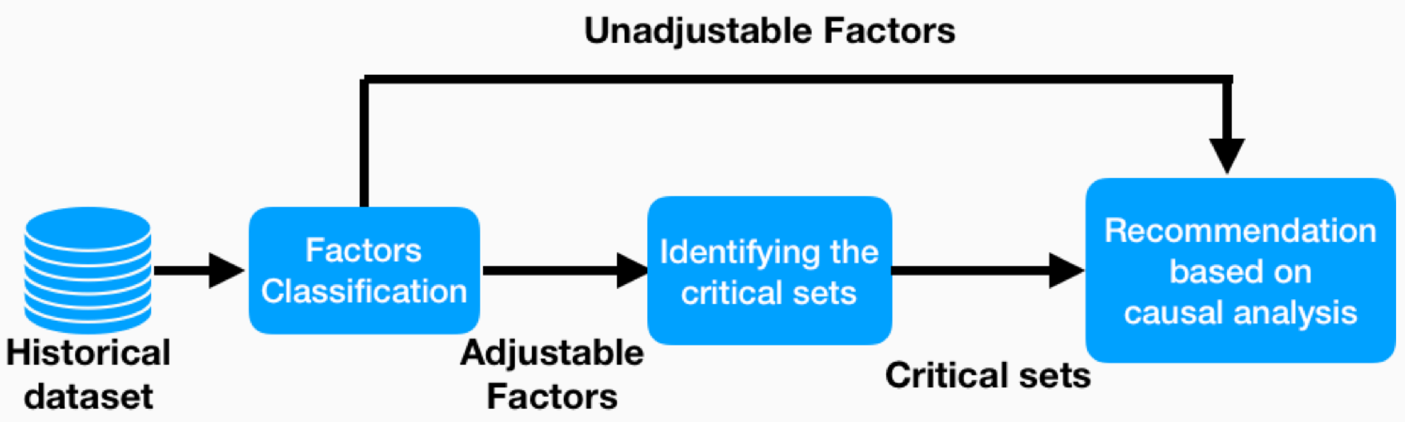


Fig. 3: The framework of *ExCause*

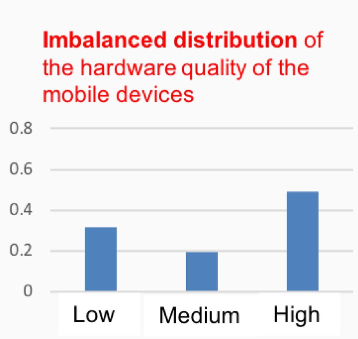
ExCause involves three stages:

1. *Factors Classification:* According to operators' experience, we classify the factors into unadjustable factors and adjustable factors. The adjustable context factors can be used to make reasonable recommendations to players. In the meantime, we should consider the impact of unadjustable context factors when we try to locate the causal adjustable context factor.
2. *Identifying the critical sets:* We identify the combination of adjustable context factors, *e.g.,* {TAN = 4G, IQ = HIGH, PD = XHDPI, OSV = iOS 12, ISP = China Mobile}, which have high ratio of unsatisfying experience, as the critical sets.
3. *Recommendation based on causal analysis:* We locate the causal context factors for critical sets using **propensity score weighting** and make recommendation for the improvements of the user experience via the historical data.

Why causal analysis?

A simple case to reveal the drawback of correlation-based method.

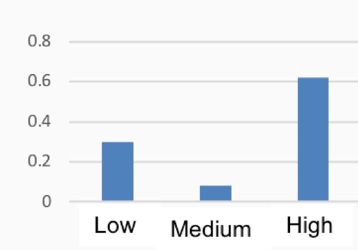
ISP	Access Network Type	Image Quality	OS Version	Pixel Density
China Union	3G	Low	Android 8.0	XHDPI



Mean(LRC)
1.527

?

ISP	Access Network Type	Image Quality	OS Version	Pixel Density
China Union	WiFi	Low	Android 8.0	XHDPI



Mean(LRC)
1.079

Fig. 4: A case based on the real dataset

Based on the real dataset, the mean of the LRC decreases, after we change the access network type from 3G to 4G. But we still can't ensure the change of the LRC is caused by the access network type, because of the imbalanced distribution of the hard quality of the mobile devices, which can also impact the LRC.

Acknowledgement

We thank the anonymous reviewers for their valuable feedbacks. We thank Juexing Liao and Ruchun Wu for their helpful suggestions and proofreading.

Recommendation via propensity score weighting

Confounder

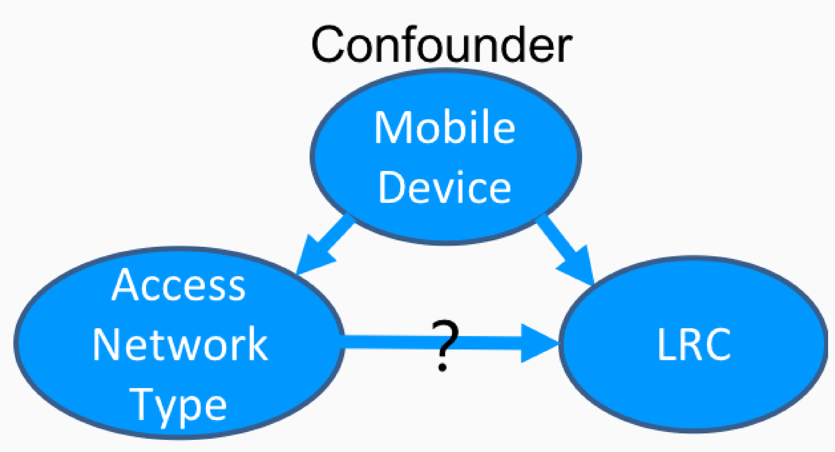


Fig. 5: Confounder for Fig. 4

In the case at Fig. 4, we intend to assess how the access network type will **cause** the change the LRC. But the quality of the mobile device impact the LRC and has imbalanced distributions of the two groups were compared. It lead to the quality of the mobile device as the **confounder** in the causal analysis context. In the paper, we apply the propensity score weighting method, which has been widely applied in social and biomedical sciences[1, 2], to deal with the confounders for two main reasons:

- conducting a controlled experiment at scale is prohibitively hard and expensive, or even impossible.
- the propensity score weighting can be used on the historical dataset.

Propensity score weighting

Propensity score:

$$e(X_i) = P(Z_i = 1|X_i) \tag{1}$$

Comparison result:

$$\widehat{ATC} = \frac{\sum_{i=1}^{N} Z_i Y_i (1 - e(X_i))/e(X_i)}{\sum_{i=1}^{N} Z_i (1 - e(X_i))/e(X_i)} - \frac{\sum_{i=1}^{N} (1 - Z_i) Y_i}{\sum_{i=1}^{N} (1 - Z_i)} \tag{2}$$

X is the confounder, which is the unadjustable factors in our paper. Y is the result in the causal analysis, which is LRC and AQ in our paper. Z is the cause in the causal analysis, which indicate the different value for the adjustable factors. The ATC is the comparison result for the two groups, which help us to determine the best adjustable factors value under different circumstances.

Result

Recommendation summary:

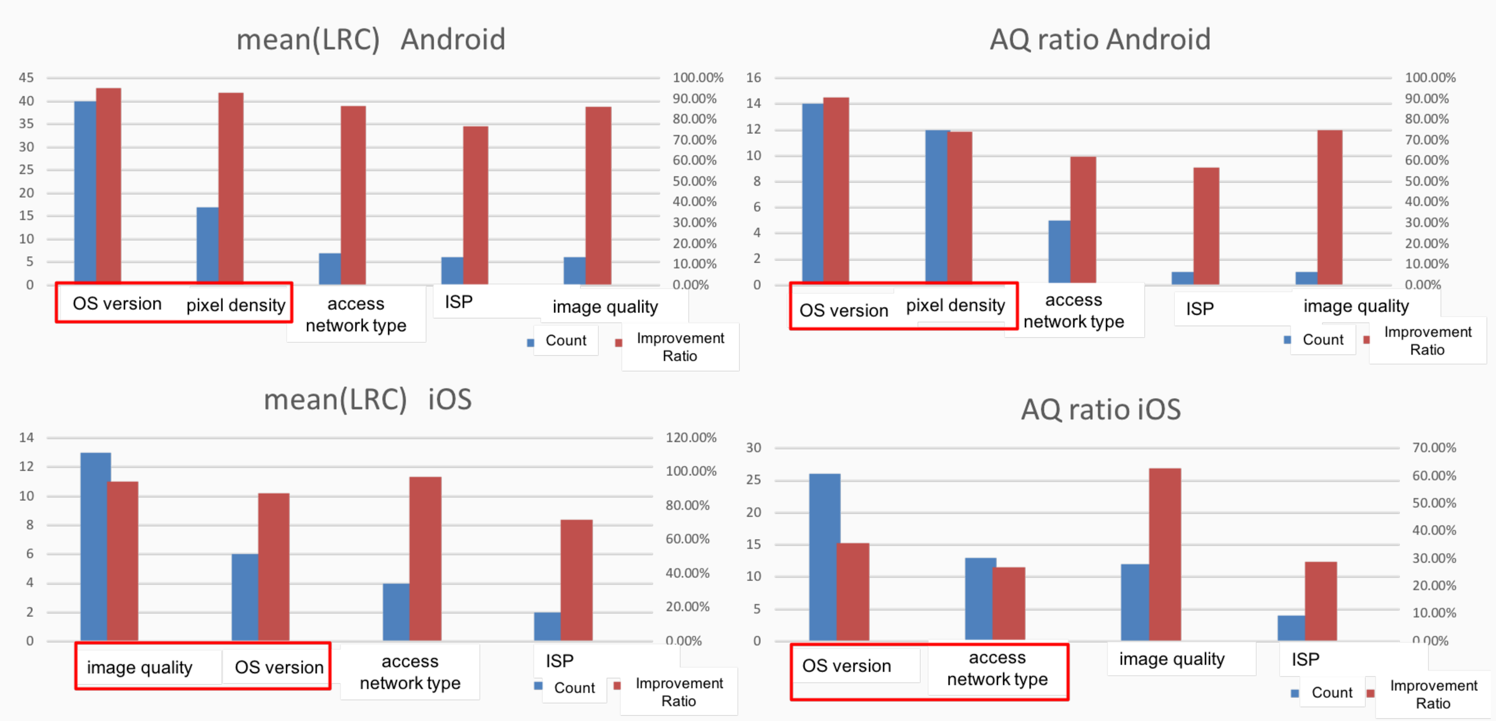


Fig. 6: Blue is the recommendation times, Red is the estimated improvement ratio

OS version, image density and pixel density as the most frequent recommended factors who need to be updated, could be the bottleneck of the unsatisfying user experience.

Benefits of causal analysis:

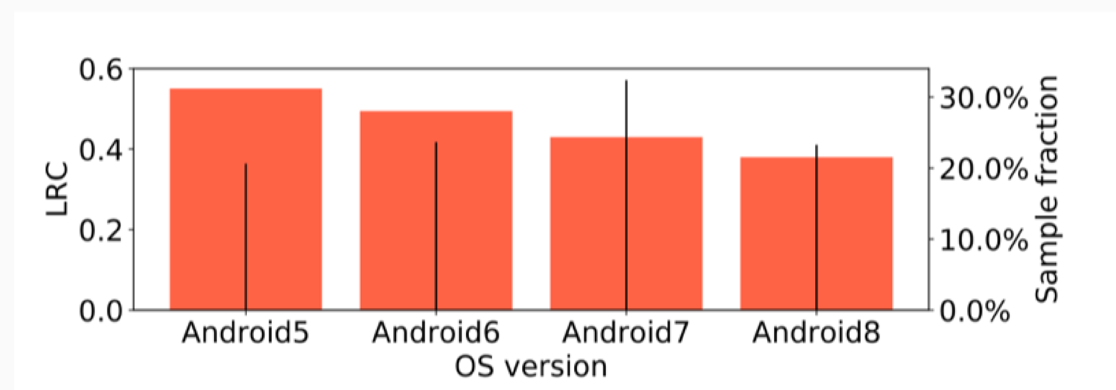


Fig. 7: The LRC for different Android OS version in the whole dataset

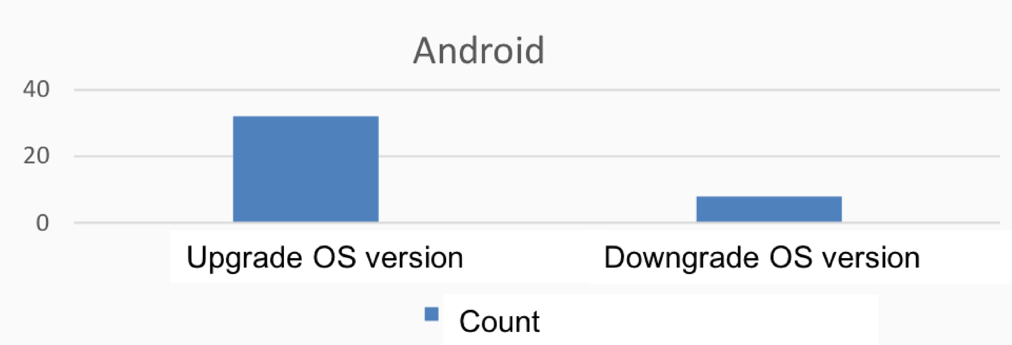


Fig. 8: Recommendation for the Android OS version

Based on the whole dataset, the higher OS version leads to lower LRC. But we find our algorithm make the recommendation of the downgrade the OS version. Because the low-end devices is incompatible with the high OS version sometimes.

References

[1] Younathan Abdia, K. B. Kulasekera, Somnath Datta, Maxwell Boakye, and Maiying Kong, "Propensity scores based methods for estimating average treatment effect and average treatment effect among treated: A comparative study: Propensity scores based methods for estimating average treatment," vol. 59, no. 5, pp. 967–985.

[2] Peter C. Austin, "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies," vol. 46, no. 3, pp. 399–424.