Clicks and Page Visits in TestSearchSatisfaction2 Schema

Mikhail Popov (Analysis & Report)
Erik Bernhardson (Engineering)
Trey Jones (Review)

May 9, 2016

Executive Summary

The original goal of this analysis was to validate the click events against the visitPage events in the TestSearchSatisfaction2 schema to see whether we can trust the click events for calculating the clickthrough rate (CTR). What we found is that we have substantially more clicks than page visits, potentially meaning that a lot of the users open results and then close them before the page has even finished loading.

Without a reliable basis to compare against, we decided to try matching the click events to the visitPage events by looking at the results position field. 9% of the sessions had click events and visitPage events that fully matched, and we used those together with sessions that had 0 click events and 0 visitPage events to get a CTR estimate. We found that the clickthrough rate estimated via the matching method resembled the naive CTR defined as ”did the user have a valid click at any point in their session.” We suspect we can use the naive CTR to run the TextCat A/B test.

As we move forward with TextCat for detecting the language of users' search queries, we want to be able to access its success (or failure) with providing users with results from Wikipedias in languages other than the one they are searching on (see TextCat A/B test). The challenge with this is that the Event Logging (EL) system as it currently exists on Wikimedia Foundation's projects was not designed nor built with inter-wiki support in mind.

Our current method of estimating the clickthrough rate uses the *visitPage* events sent by the user's browser when they have opened one of the search results and the page has loaded, but it only works within the same wiki and would not work if they opened a result that was on a different wiki. To that end, we implemented the *click* events in our schema as a way to capture clickthroughs from the search engine results page (SERP) directly, so that we could have a sense of users' engagement with the search results where we cannot track their page visits. In this report we investigate the validity of the *click* events for calculating the clickthrough rate so that we can possibly use it to assess the impact of TextCat language detection on the quality as well as the quantity of search results we return to the users.

Throughout this report, *sessions* actually means *sessions with nonzero result sets*, since we cannot expect the user to click on any result if they were not given any results.

Figure 1: Daily clickthrough rate by estimation method and platform.

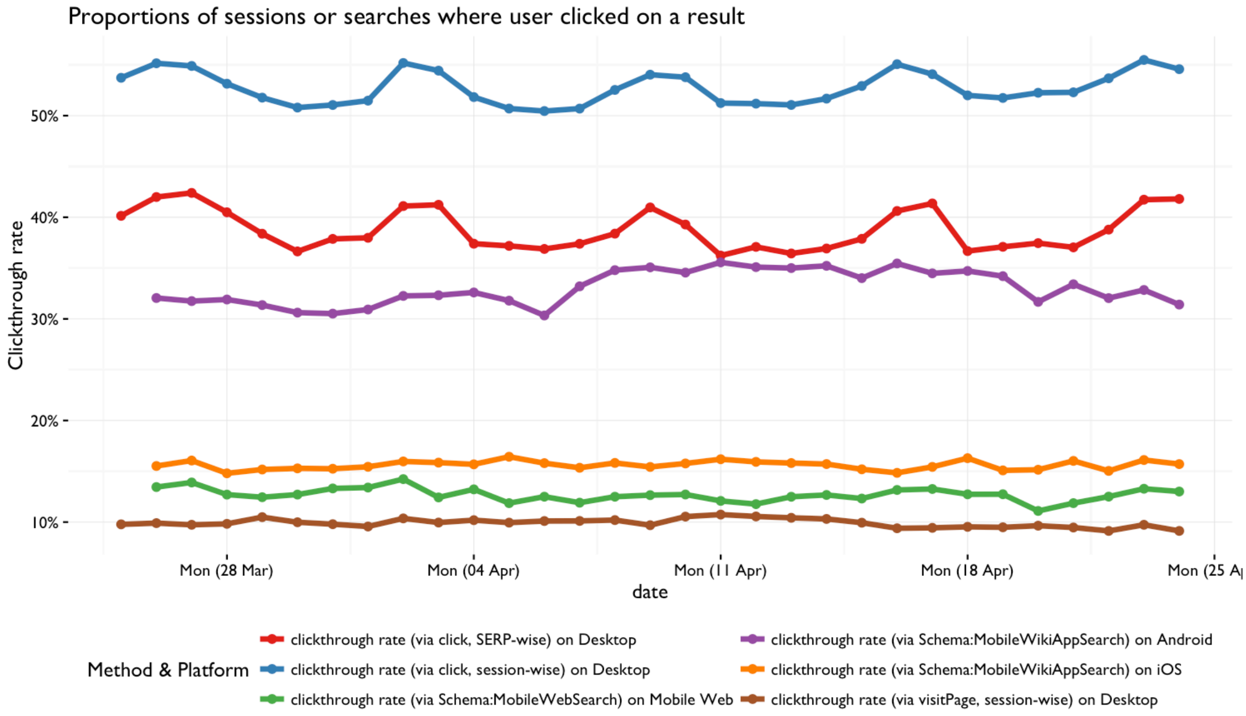

Figure 2: Daily clickthrough rate by estimation method, including matching click and visitPage events.

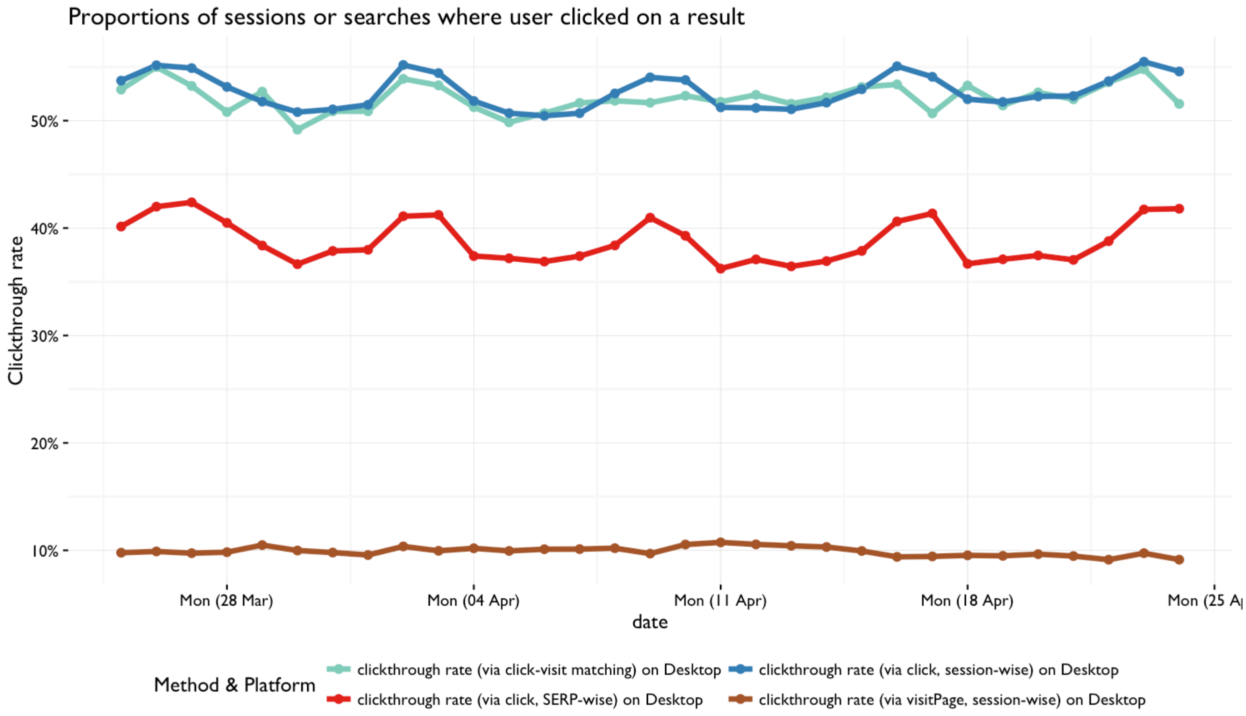

Table 1: Clicks and page visits in sessions. A "valid click" refers to ***click*** *event that had a* ***result position*** *recorded (indicating which result in the list of the returned results they clicked on), and likewise for "valid visit"s which refer to* ***visitPage*** *events.*

	Sessions	
sessions with some valid clicks	89.127%	
sessions with some valid visits	16.817%	
sessions with some valid clicks AND some valid visits	14.237%	
sessions with more valid clicks than valid visits	3.378%	(23.730% of 14.237%)
sessions with more valid visits than valid clicks	0.862%	(6.052% of 14.237%)
sessions with valid clicks AND valid visits, AND clicks match visits 100%	8.995%	(63.183% of 14.237%)
sessions with valid clicks AND valid visits, BUT clicks don't match visits at all	0.743%	(5.216% of 14.237%)
sessions with valid clicks that couldn't be matched with valid visits	4.538%	(31.878% of 14.237%)
sessions with valid visits that couldn't be matched with valid clicks	2.711%	(19.040% of 14.237%)

Table 2: Overall clickthrough rate, estimated with different methods.

Method	**Clickthrough Rate (%)**
via click-visit matching	52.031
via visitPage, session-wise	9.979
via click, session-wise	52.575
via click, SERP-wise	38.571

We were intrigued by the idea of sessions where the *click* events perfectly matched the *visitPage* events, rather than sessions where there were more *click* events than *visitPage* events (meaning the user clicked on a result but then closed the new tab before the page even loaded). So we narrowed our gaze at sessions that had either 0 clicks/visits (abandoned searches) or 100% matched clicks and visits. We suspect this may be a more accurate way to estimate clickthrough rate, albeit impossible if *visitPage* events are not available, such as in the case of the TextCat A/B test. Surprisingly, the resulting clickthrough rate is really close to the clickthrough rate when looking at *click* events per-session in Figure 2!

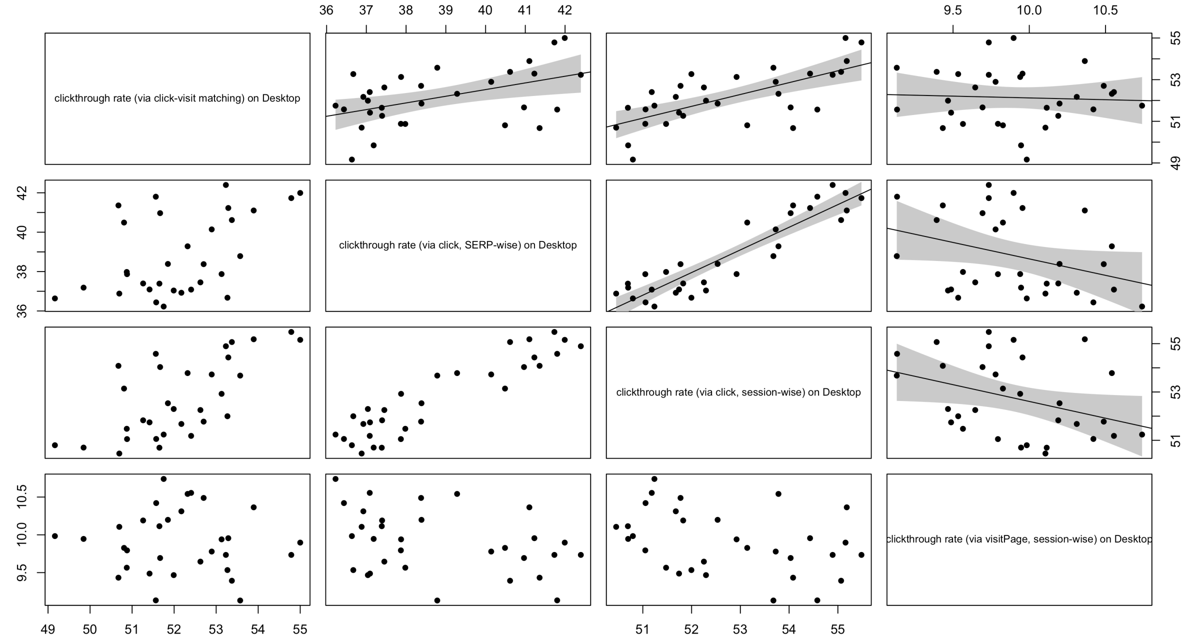

Figure 3: We also looked at the correlation between the two time series and did not see evidence of a strong, positive linear relationship between CTR-via-matching (Row 1) and CTR-via-session-wise-click (Column 3), and instead saw the clickthrough rates scattered too much. Two identical time series would have a perfect positive (1:1) linear relationship.

We performed the Granger causality test to see if the clickthrough rate we will be able to estimate can be used to reliably forecast the clickthrough rate that will not be able to estimate (due to absence of *visitPage* events). Since the Granger causality test does not test zero-lag (immediate) causality, we shifted the time series so the CTR-via-matching is on a 1 day delay from CTR-via-session-wise-click, so correlation between the two would translate to forecastability. We found that we cannot use the latter to forecast the former. So we cannot **reliably** use one estimated clickthrough rate to estimate another.

We suspect that in the absence of *visitPage* events, we can use a combination of the two other, *click*-reliant clickthrough rates (one calculated on a per-session basis, the other on a per-SERP basis) to estimate the clickthrough rate as it would have looked if we had *visitPage* events available to match with *click* events. **Alternatively** – and this is the simpler option, albeit a less accurate one as mentioned above – we can use just the overall (rather than daily) session-wise clickthrough rate (see Table 2) as a very rough proxy/approximation for the more accurate, but non-observable overall clickthrough rate, either as it is or by multiplying it by some constant.