Don't Click Here

**David A. Shamma
and Malcolm Slaney**
Yahoo! Research

Everybody was amazed a few years ago. How could Amazon patent one click? Isn't that obvious? Haven't we been single and double clicking for some time? The US Patent and Trademark office decides what is *patentable*, which they define as things that are new, useful, and nonobvious. If that all checks out, the patent office clicks "approve."

Since then, the landscape has changed considerably. Single clicks, touches, and gestures are now powerful, often complex interactions that are part of our everyday vernacular. Currently, touchless interactions and zero-click interfaces bring us content, push notifications, and recommended playlists (see the sidebar for definitions of the various click-data measures). Beyond this, the absence of interaction, the minus one click, is taking the forefront. But more about that shortly.

We multimedia people live in a world of data. The large Internet companies know more about our viewing and purchasing habits then even our mothers do (or should). An important reason for their success is personalized services; companies retain purchase history and use that information to recommend new content and provide new products. Netflix wouldn't be as successful with just a catalog of videos and a search engine.

This means that one of the most important aspects of multimedia research is mining data about user behavior. Obviously, the more data about people you can collect, the better decisions you can make. In a world with one or fewer clicks per piece of content, one needs a lot of users to get a meaningful signal. Cutting-edge, interactive-multimedia work requires thousands to millions of people, so a lab study with 20 users for three weeks doesn't begin to cut it.

However, the bigger problem is one of instrumentation: What do we measure, and why do we collect it? More often than not, existing instrumentation is not geared toward finding engagement. Tim Berners-Lee recently observed that data is collected to benefit the people buying consumer data, not the consumers that produce the data in the first place.[1] Computer science research tends to address people (users) but not customers—you know, the marketers and advertisers paying for our "free" video hosting and email—except perhaps in the emerging computational-advertising field. A monetization ecosystem now directs multimedia research and is changing how we think about people, interaction, media engagement, and data instrumentation. And it all started with one click.

One-Click Interfaces

What makes the Amazon idea of a one-click store so amazing is the hidden process. Before the one-click idea, purchasing something meant a click to put something in your shopping cart, a second click to view the cart, a third click to checkout, and then innumerable clicks and data entry to complete the order. The result of all these required interactions was a high rate of shopping-cart abandonment—something no merchant wants to see. Of course,

Click-Data Measures

Various multimedia experiences require different kinds of click or touch interactions. Most importantly, the absence of clicks should be disambiguated when measuring engagement.

- *Many clicks*. Example application interactions might include online video editing.
- *One click*. Providers can roll several actions into one clear action, such as one-click purchasing.
- *Zero click*. This involves collecting information as a side effect of other interactions, such as measuring click-through as a function of (random) descriptive image shown.
- *Minus one click*. The user is provided a shortcut but opts not to take it— for example, by not choosing "Click here to skip this."

now it all seems obvious. One click to buy anything.

Click data drives many of the best Internet products and remains the primary Web metric. With enough data about what people click on, search engines can return only the most popular results. For a time, a Web image search for the term "Christmas" presented a wonderful array of classic Christmas pictures and one picture of a pretty woman in a red and white bikini on her Christmas card. You can guess which image got lots of clicks and thus stayed on the first page. Similarly, as discussed in a previous Visions and Views article,[2] YouTube videos often use a salacious thumbnail to get people's attention.

More subtle ranking approaches consider click data as just one of many signals about the potential search results. PageRank was one of the best ways to rank webpages in the early days of Web searching. The current method, a paradigm known as "learning to rank," uses data in massive search logs to learn the proper combination of features to build highly effective search engines.[3] These features include information such as "Does the page's title include the search query?" or the distance between the query and the page in a model of lexical topic space. Such features plus the click data are quite successful at modeling which links get clicks.

Zero-Click Interfaces

So, click data is useful. Lately though, product managers have been asking for zero-click interfaces. One might wonder how anything can be done with no clicks? The answer is nothing, by itself. So how do we get useful information? Information in a zero-click interface comes as a byproduct of other interactions.

For examples, colleagues at Yahoo built a system to find the best images to go with a concept. One might use this information to illustrate a restaurant on the search-results page. Statistically speaking, people are more likely to click on a search result if there is an image as part of the summary. Thus, search engines try to add relevant images and other content to each search result to make it more attractive.

But which image is best? We could ask people, but more often, the preferred approach is to just use Web-scale data. A zero-click interface discovers the best image by putting a "random"
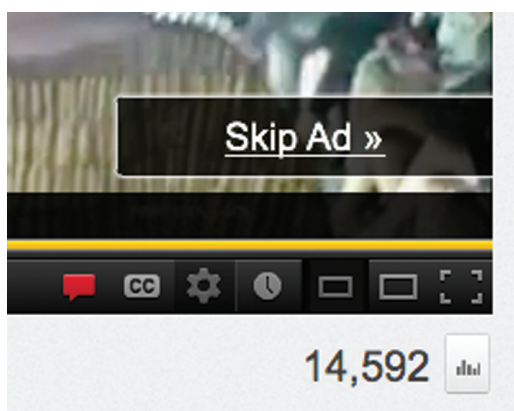

Figure 1. Minus-one-click interface. YouTube plays an ad before a video and adds a button that allows users to skip it. Advertisers only get charged if a user watches the ad in full.

(but good) image in front of each user. Users go about their business, searching for restaurants, say, and clicking on the ones that best match their needs. However, some of the collateral images are better than others. As a byproduct of the search page, we get the information we need to help us to display the best image for a restaurant in many different contexts.

Minus-One-Click Interfaces

Believe it or not, even zero clicks is sometimes too much to ask from a user. Lately, Hulu, YouTube, and other video sites have realized this. Thus, as a method of measuring engagement and advertising efficacy, they have started paying advertisers when users who do *not* click on their ad. What? Wait a minute, I thought we wanted people to click on our ads. Isn't that good? Yes, but sometimes it is even better if they do not.

The hard truth is that it is almost impossible to measure engagement with multimedia content without recording action. If a movie, a TV show, or even an advertisement is engaging, it holds our attention and we sit there watching it, held in suspense, laughter, or fear. As in other social contexts, that is not the time to act (click) on something.[4] In this instance, if you click on something, we can infer you are, in fact, not engaged with the content.

In 2010, YouTube rolled out the minus-one-click interface by displaying a pre-roll ad for a few seconds and then popping up a rather large button that says "Click here to skip this ad"[5] for the remainder of the 30 second spot (see Figure 1). If you click it and skip the ad,

> ## It is almost impossible to measure engagement with multimedia content without recording action.

the advertiser does not get charged for the ad. Recently, Hulu decided to do this without any click or button at all.[6] That is, it will only charge advertisers when users do not click away and view an ad in full. If you close the window or browse to other content 25 seconds into that 30 second spot, the advertiser does not get charged. Let us frame this in a different way. You pay $12.50 to go see a 90-minute movie. If you leave 83 minutes into the movie, you get your money back.

As an advertiser, knowing people did not leave before the end of an ad provides value; they know their message was effective and seen in full. Advertisers will, and do, pay a higher premium for documentation that people didn't lose interest in their story. The minus-one-click interface actually asks you not to do something, sometimes by displaying a button that beckons you to click it with the hope that you don't. The obligation for action is on the advertiser. They must make a good ad, one that will keep you from doing anything and will keep you engaged. People who do not like the ad can interact and get to their desired content faster. Internet sites have begun giving people the choice and report this choice back to the paying advertisers.

New Signals and Multimedia Research

In general, it is hard to get meaningful user data. It's even harder with multimedia applications and experiences, which people often experience by huddling around a cell phone or eight-feet away from the TV on the wall. To date, multimedia systems generally deal with explicit signals, such as requesting content, star ratings, and skipping and reversing.[7] We can no longer rely on explicit-interaction data alone. Research on implicit signals that we see in other domains,[8] how people interact with each other and with content, is just now entering the multimedia picture. The onset of the minus-one-click interface is an example of measuring engagement by the lack of signal—the empty space in the dataset. Although counting clicks data is still the most common means of instrumentation, sometimes the absence of a click is even more valuable for understanding engagement and increasing monetization.

Think about these information sources and how much information you are requesting from the user as you design your next multimedia experience. Some days you will be lucky to get the right click data, but really you should also design and instrument for a minus-one click. **MM**

References

1. I. Katz, ''Tim Berners-Lee: Demand Your Data from Google and Facebook,'' *The Guardian,* 18 Apr. 2012; www.guardian.co.uk/technology/2012/apr/18/tim-berners-lee-google-facebook.

2. M. Slaney, ''Precision-Recall Is Wrong for Multimedia,'' *IEEE Multimedia,* vol. 18, no. 3, 2011, pp. 4–7.

3. T.-Y. Liu, ''Learning to Rank for Information Retrieval. *Foundations and Trends in Information Retrieval,* vol. 3, no. 3, 2009, pp. 225–331.

4. D.A. Shamma and Y. Liu, ''Zync with Me: Synchronized Sharing of Video through Instant Messaging,'' *Social Interactive Television: Immersive Shared Experiences and Perspectives, Information Science Reference,* IGI Global, 2009.

5. M. Wilson, ''YouTube Gets It, Will Allow Ad Skipping,'' blog, Gizmodo, 30 June 2010; http://gizmodo.com/5576453/youtube-gets-it-will-allow-ad-skipping.

6. W. Weaton, ''Google Is Making a Huge and Annoying Mistake,'' blog, Hulu, 2 May 2012; http://blog.hulu.com/2012/04/17/100.

7. T. Syeda-Mahmood and D. Ponceleon, ''Learning Video Browsing Behavior and its Application in the Generation of Video Previews,'' *Proc. 9th ACM Int'l Conf. Multimedia,* ACM Press, 2001, pp. 119–128.

8. A. Pentland, *Honest Signals: How They Shape Our World,* MIT Press, 2008.

David A. Shamma is research scientist in the Internet Experiences Group at Yahoo! Research. Contact him at aymans@acm.org.

Malcolm Slaney is a principal scientist at Yahoo! Research and a consulting professor at the Stanford Center for Computer Research in Music and Acoustics. Contact him at malcolm@ieee.org.