Introduction

Our technology focuses on using user behavior data combined with the characteristics of search results to find the best possible match between users and content. User behavior data is also valuable in demoting or excluding search results that are poor matches or undesirable.

As early as 2013, we began working with adult media libraries. These libraries have different characteristics from mainstream libraries of music videos or how-to videos or educational lectures that we had encountered before.

Without being lewd, the reasons a thirty-year-old male viewer might be finished with a video in an adult media library after five minutes of its twenty-minute runtime are diverse; it could signal high quality, low quality, or a spouse who arrived home from work early.

This is where the application of our technology to an adult media library began, with some key initial insights from reviewing existing behavioral data and talking with platforms:

- Events – including news events – can substantially affect the viewership of this type of content.
- Demographics are more important in the adult search use case, including gender, age, location, and so on.
- Abandoned or partial viewings of media – particularly video – must be interpreted carefully.
- People viewing adult photos and videos expect a seamless experience that includes query and non-query search.

There are two different ways for a user's media session to end and we capture both of these: termination and abandonment. Termination is the term we use to describe the end of a session, where a user completely leaves the ecosystem of a media library. Abandonment, meanwhile, is how we describe a user's departure from a media item in favor of another media item.

We segregate terminations from departures and treat them differently in the context of user behavior analysis. Across large datasets, a single library may experience millions of terminations and hundreds of millions of abandonments. Beyond this, there are often automated views and abandonments (and terminations) driven by botnet traffic or advertising efforts to exaggerate the popularity of certain items.

The reason this segregation and classification is so important is that people depart media items prematurely for a variety of reasons. Because termination is a more "severe" event from an advertising perspective – since the user may leave the advertising audience completely – the media items driving a large number of terminations may require special attention.[1]

In cases of membership-model adult sites, users can be tracked at high levels of granularity and identity continuity, since they are logged in. Often, the user database includes a username, a real name (or at least credit card name[2]), payments information, password information, the user's search history, synthetic variables expressing a user's preferences,[3] entry point analysis,[4] and social (comments/likes/etc.).

[1] In many cases, these items may be ones with misleading titles, or containing problematic advertising, and so on.

[2] About twenty percent of adult site members pay with company credit cards, prepaid debit cards, or other methods where their identities are not clear from the payment method.

[3] This may include favorite or recurring search terms, probable user sexual orientation, and even metadata like geolocation and advertising tracking data.

[4] Includes sourcing and routing data on referrals, search engine pedigree data, and banner activity.

Demographic information can be useful and appended even if individuals are not logged in or are unidentified. Because individuals often access sites from the same or similar devices and addresses, and with identical or extremely-similar searches, it is possible to build cohorts of users that are similar; this cohort-level analysis was the focus of much of our early research.

- Identifying the preferences of (and optimal results for) non-logged-in users is important as these are potential members for membership sites.
- On many adult platforms, non-logged-in users are the majority. Hence, targeting the right mix of library content and advertising for these unauthenticated users is more than a "nice to have," it's a user experience issue.
- One of the areas where adult platforms compete is load time and video performance; the faster the right result can be served to the user, the shorter the total time between query and load.

Haystack's core technology[5] is a set of algorithms that locates and assembles similar cohorts unauthenticated users or cohorts of users. By noticing the first user's behavior and serving results that user favors to a subsequent similar user, our technology gives all users in the cohort better results. This increases the appeal of the platform, the effectiveness of content and advertising, and user trust in the result set.

We are always doing research to improve our algorithmic work and much of this is empirical research studying natural experiments and abnormal shocks. Two recent events in Q417 and Q118 gave us opportunities to study user behavior at volume, including large abandonment and termination events.

Location Research Initiative

Contextual behavior is important, which includes the importation of non-platform data. This includes seasonal/predictable items (the timing of the Superbowl, for instance) as well as unanticipated or surprise events.

In a recent "surprise event" scenario, at 8:07am (local time) on January 13, 2018, an erroneous alert of an incoming ballistic missile was broadcast via the cellular network in Hawaii. By 8:23am, viewership on the adult media library in question had dropped 77%[6] to 90%[7] across various metrics. Search traffic (query execution) during that period was down 88%.

For context, typical variation in regional viewership and search is between 8% and 10% intraday, with predictable decreases while people are commuting, at lunch, and so forth. Decreases of more than 30% in viewership or more than 40% in search are highly unusual, and decreases larger than 50% are very rare.

This seems to indicate that – barring life-threatening disasters – the level of consumption of pornography is remarkably stable and robust against all but the most extreme shocks. Interestingly, during the panic, some people were searching for items and leaving browsers active (particularly on mobile), perhaps suggesting they planned to watch results later once they reached safety.

It seems many people in Hawaii, when (erroneously) alerted to the incoming missile at 8:07am, managed to stop watching adult content, gather their belongings or pets, and get approximately fifteen minutes from home. The "false alarm" statement was issued at 8:45am and at exactly 8:59am traffic and search levels both returned to normal. By 9:01am, traffic levels

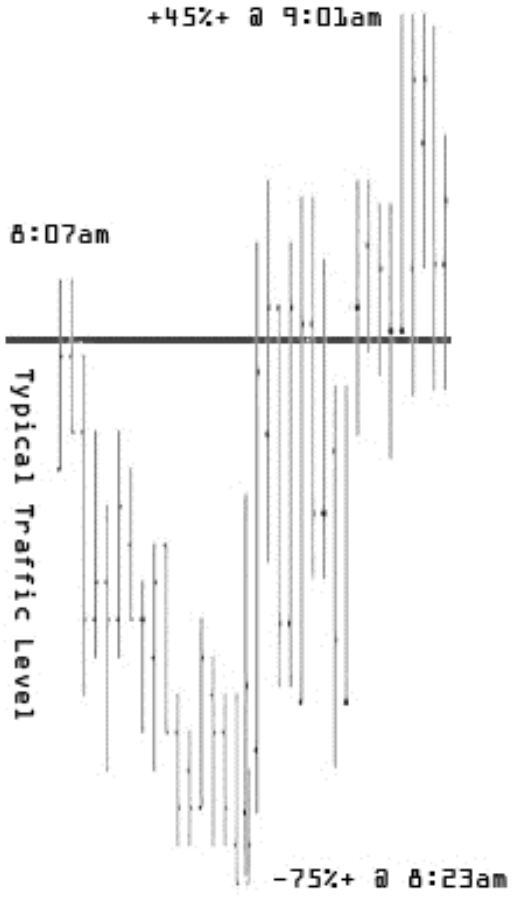
[5] Core technology, in this context, refers to technologies covered by U.S. Patents 9,262,526 (2016) and 9,594,809 (2017) and research done on successor technologies.

[6] As compared to average activity from Hawaiian viewers on the previous ten Saturdays between 8am and 9am local time.

[7] As compared to change-in-viewership during the same period in London at the same time (6:23pm).

were 48% higher than normal for 9:01am on a Saturday.

There was an unintended result of this light traffic period, which was that some searches were "overweighted" at 9:01am due to a suddenly-thin record of searches. For instance, actress Penny Pax was not among the top 40 searches on the platform prior to the missile warning decline, but became the 22nd-most-popular search during this lull in activity.



Due to this thin volume and substantial viewing time on her videos, Penny Pax's ranking was amplified for hours in search rankings on the platform. This issue was studied and resolved within a week using several features, including the quarantine of user behavior during these unusual usage patterns to prevent overweighting of results that are rarely-searched and atypical.

As a result of these revisions, atypical or overly-niche results are unlikely to present as recommended media items or promoted results for more than one minute, or four cycles of the algorithm's re-tuning.[8] Our hope is to continue to improve this optimization and for unusual presentations of overweighted niche content to be largely eliminated in our next release.

Gender Research Initiative

In January of 2018, well-documented "Women's March" events occurred in multiple cities across America (but not simultaneously in most other countries). The attendance at these events was estimated by both regional and national media outlets using data from event organizers, aerial observation from drones and news helicopters, event licenses filed with various municipalities, and local police forces' crowd control estimates.

This presents an interesting opportunity to study the impact of a large mostly-female event on the female viewing population of an adult video archive. Around 10:40am local time in Chicago (one of the larger Women's March events), traffic fell substantially (as did search activity). This trend continued for several hours.

For users flagged as female in the database (either through demographic generalization or through membership data), viewership fell as much as 6% in the afternoon and close to 9% by the late afternoon. Research using credit card given names for gender sorting suggests female viewership dropped by slightly less than this, but these numbers may be distorted.[9]

[8] The algorithm re-tunes and revises search prioritizations every 250 milliseconds.

[9] It is more common for married heterosexual couples to use credit cards issued in the male partner's name. It is also common for name-gender filters to overclassify names like Sam or Chris as male (when they could stand for Samantha or Christine).

Interestingly, in cities where large women's marches occurred, "porn for women" was a trending (and even #1 in six cities) search term within this database following the event.

Comparing to asynchronous data[10] from Hong Kong (which normally exhibits similar female adult content viewership characteristics to Chicago), where no significant attendance at a women-centric protest event was recorded, we see significant decreases. Similar decreases are observed in comparisons against previous non-Women's-March weekend times.

This research suggested tracking events that may compete for attention with adult content for specific demographics is important. A retrospective audit also finds smaller decreases in regional female viewership during San Francisco's Pride Parade and even a (small) detectable decrease during the WNBA Finals.

Haystack's technology – revised as a result of this research – allows a "temporal quarantine" where viewership and search activity for certain groups during these unusual events is flagged as atypical and not used to influence (and possibly unduly-skew) search results.

Privacy Concerns

There are special privacy concerns for adult content and adult search user activity records.

Obviously, fully anonymizing user records can be disruptive to search optimization efforts but may give users a degree of comfort. The solution is a cohort layer, which our technology creates, where user data is generally analyzed in anonymized blocks but can be resolved down to individual user activity when needed.

During Q417, a series of media reports emerged that contained sexual harassment allegations against various celebrities and executives. It might be relevant to these discussions if a celebrity in question had an account on the adult media platform or had a tendency to search for (or receive optimized results containing) rape-themed pornography, for instance.

It is important to preserve the absolute secrecy of search trails, including individual search queries and "breadcrumb" series of recommendations.

Ideally, the precise characteristics of search activity would be preserved while individual user data linked to account names, credit card payment information, or other identifiable data is never expressed in unencrypted form.

Our technology is used on platforms that have different specific privacy policies, but all have substantial concerns about user privacy.

In a separate design and engineering effort, we were working on a blockchain solution where user-level query data is stored in an encrypted, secure format that could not be associated with any one user but could be united with that user's record to optimize search results.

After this series of discussions around privacy, our production of the first master data dictionary for storing search data on the blockchain was accelerated by two months, with at least one partner using the implementation in its pre-production version in February 2018. This is the basis for our secure user data storage system.

Our goal is to promulgate and continually test standards for user data storage (including behavioral data) to ensure users feel comfortable expressing their interests and preferences in search queries on adult media platforms. These standards are already being discussed and vetted by the world's largest adult media providers.

The Compilation Complication

It's common in adult content – more common than in other types of media – to have users or curators create videos that include a series of constituent videos.

[10] In other words, 10am in Chicago is compared to viewership at 10am in Hong Kong.

Often, a user or series of users will "rip" a clip[11] and combine it with another video (or videos) into a compilation. It may be thematic, containing similar adult content, or may be a montage of videos that feature the same actor or some other common thread. In some cases, audio is altered or inserted to give the impression of continuity between the clips.

This adult media library included hundreds of these compilations, which are problematic for a variety of reasons:

- They may be duplicative, containing content that is already in the library.
- They may have been created in violation of copyright laws.
- They may pollute or devalue the viewership data for the actors or studios involved, since many users only view a portion of a compilation video.

After studying over 250 compilation videos containing at least two independent pieces of adult content, we created an engine that adjusts and "zeroes" the timeline for such videos at each break in footage. In other words, a compilation containing three five-minute videos is treated that way, rather than as a fifteen-minute video.

This technology – developed as early as 2016 and still improving – is in use on other non-adult platforms, including YouTube, where compilations of music videos and other content are increasingly-common as the accessibility of video-editing software increases. We also met with UK-based search platform Shazam to discuss how its technology could be used to detect compilations and duplicate videos; this research and these discussions are ongoing.

Cross-Platform Activity Drives Traffic

The popularity of content in adult libraries changes more quickly than in other types of media libraries. A few tweets or a few posts on Instagram is sufficient to get users searching for new actors or actresses (usually actresses) or search terms across the major adult platforms.

To understand this phenomenon, we studied 20 female and 20 male adult performers whose names' search rankings changed substantially across major platforms in terms of activity.[12] In particular, we studied performers whose content became substantially more popular without paid advertising (which can skew or bias search results) or other interventions/circumventions.

One actress, in particular, caught our attention. Not only did searches for her (pseudonymous) name increase dramatically from June 2017 to December 2017, but this increase seemed linked to activity on non-video platforms, emphasizing the importance of monitoring and integrating cross-platform data. Around December 2017, despite essentially no paid advertising, Karma Rx reached one million followers on Instagram and one hundred thousand followers on Twitter.

Wanting to better-understand how this was achieved, we formally brought L.A.-based Karma Rx on as an advisor to Haystack in December of 2017, with our CEO making two trips to Los Angeles in January 2018 and meeting with key people as well as startups doing cross-platform optimization and social network analytics work.

Like many up-and-coming adult social media stars, Karma Rx managed her own accounts and had no advertising budget. Much of her success was related to her posting of provocative, timely content on multiple platforms, all of which eventually drove users toward paid content. This led to a staggering 2000% increase in her followers on Instagram and a huge increase in searches for her name on a major adult site.[13]

Karma Rx's success story became not only a source of knowledge about the industry, but an example of how cross-platform promotion could change search behavior almost instantly.

[11] This may be a legal or illegal download.
[12] This is a proprietary calculation.
[13] Figures calculated over an 18-month period.

Karma Rx's cross-platform presence extends beyond her online content. With retweets and likes, she substantially raised Haystack's level of visibility at AVN 2018, an important adult industry event in Las Vegas where key adult sites and other potential customers were present.

Studying the success of individual stars like Karma Rx lets Haystack understand how the popularity of these people begins, spreads, and persists across platforms from adult video sites to mainstream apps like Instagram and Snapchat.

Karma Rx.

Monitoring Cross-Platform Promotion

After studying the rise in Karma Rx's popularity, we brought aboard several advisors with a deep understanding of how social media platforms drive viewership activity. These advisors included Brooke A.S. Ricketts, former Head of Brand Strategy at Twitter. We also brought aboard several people with experience at the content platforms themselves, including Feven Woldu, a former Pandora executive, and Dayo Olopade, a current YouTube executive (and former Facebook executive).

We also interviewed key early developers, technical staff, and marketing/promotions individuals from top adult sites, including personnel focused on search engine design, search engine optimization, and search result tuning/improvement. These discussions included interviews of past employees of Mindgeek subsidiaries Brazzers and Pornhub and of WGCZ Holding venture XVideos.[14]

A well-timed tweet from Karma Rx (with over 100,000 followers on Twitter) could drive up to 1,500 supernormal video views within half an hour; a high level of audience-attentiveness and a very high response to broadcast (as opposed to narrow-cast) marketing. This suggests cross-platform algorithms – including our own – may underestimate the performance of tweets, Instagram stories, and other cross-platform content from particularly popular personalities (or those with particularly loyal followings).

Ongoing Research

The team continues to do diagnostic semi-structured interviews with top content curators and creators, including performers like Karma Rx and channel-level stars like Matt Farah.[15] The goal of this research is to isolate which approaches to handling searches amplify (or defeat) efforts to promote content, with a particular focus on search and sort mechanisms that may work counterintuitively or problematically.

[14] XVideos streams close to 4TB per second of adult videos, much of it in HD or better quality (2018).

[15] Matt has hundreds of millions of views on his YouTube channel TheSmokingTire and is an advisor to Haystack.

The team's work to develop core technology in the lineage of U.S. Pat. App. 15/435,077 and other research[16] is ongoing.

In 2018, the primary research areas are three:

- Where can we apply video learnings to audio, music learnings to podcast content, and so on?
- To what extent, according to which rules, and with what level of anticipated lag does social media activity drive search activity or library access?
- Are some news stories, superstar music videos, and so on different enough from the rest of the content in a library to justify totally separate search rules?

As this research progresses, we will continue to present our findings and theories at major universities, in the developer community, and with key partners who are able to inform our work and our research paths.

Conclusion

Adult video libraries are not only an exciting application of our technology but also an example of how each type of media requires a specialized approach to audiences, content, metrics, and search. Each month, we work with more adult media sites to educate them about the advantages of strong recommendation and search tools – both for members and for increasing the conversion rate of non-members.

Unique research opportunities abound in the adult video space, in part because certain assumptions can be made about usage patterns – that people are not huddled together watching porn at the Women's March, for instance. These allow us to study viewership patterns, audience interests, and the effects of outside events and stimuli in a way that is more difficult when examining other kinds of media.

Adult platforms using our data have enormously robust data archives and relatively good records of site version control. This allows both A-B testing (where different engine versions or different site instances are available simultaneously to cohorts of users) and longitudinal assessment (where either content library or site design changes over time affect the user experience or to which records a user might be referred with similar search behavior).

The adult industry is a leader in the adoption and deployment of new technologies. This was true in VHS cassettes, pay-per-view programming, and in-room hotel video. Adult content sites were among the first to move to 1080p and now are among the first to have substantial content libraries in 4K. We've spoken to multiple adult content libraries building content inventories in AR, VR, and other alternative formats. Because adult content sites are often early adopters of new technology, we see them as a "leading edge" opportunity to understand search tools that more mainstream sites may later require.

We hope this rapidly-expanding area of multimedia search continues to benefit from Haystack technologies and continues to educate our team about its audiences, special considerations, and use cases. Our work with adult content will continue and has already resulted in innovations that benefit "mainstream" customers, including non-video customers like podcast platforms and music library providers.

Contact

To contact the Haystack team responsible for this case study, email us:

haystack@weapons.systems

For the latest on Haystack, follow us on Twitter:

www.twitter.com/haystack

[16] This research is protected by a mix of patents, pending patents, and trade secrets.