

# Technical Disclosure Commons

---

Defensive Publications Series

---

May 19, 2016

## VIDEO RECOMMENDATION ALGORITHM

Scott Davies

Justin Lewis

Follow this and additional works at: [http://www.tdcommons.org/dpubs\\_series](http://www.tdcommons.org/dpubs_series)

---

### Recommended Citation

Davies, Scott and Lewis, Justin, "VIDEO RECOMMENDATION ALGORITHM", Technical Disclosure Commons, (May 19, 2016)  
[http://www.tdcommons.org/dpubs\\_series/200](http://www.tdcommons.org/dpubs_series/200)



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

## VIDEO RECOMMENDATION ALGORITHM

Users on social media sites, such as YouTube, may have dramatically different viewing behaviors, based on a number of factors. Many of these factors are considered by social media recommendation algorithms when identifying videos to recommend to a user. One commonly used algorithm is largely based on co-viewership. This algorithm relies on monitoring the videos that a user watches and identifying a group of other users that have also watched the same or similar video genres. The videos that have been watched by other members of the group, but not watched by the original user, may be selected as a recommendation for the original user with the assumption that the users in the group all have similar interests. This approach generally results in video recommendations that align well with the original users' interests.

There are some additional factors which may not be taken into account by the co-viewership models. For example, videos are often watched by different users for different reasons, and determining the user's motivations for watching a video may be difficult. The majority of watch time on a social media service may come from content published by established creators, but the majority of session starts come from casual viewers. Videos are often polarized and tend to fall into one of two categories.

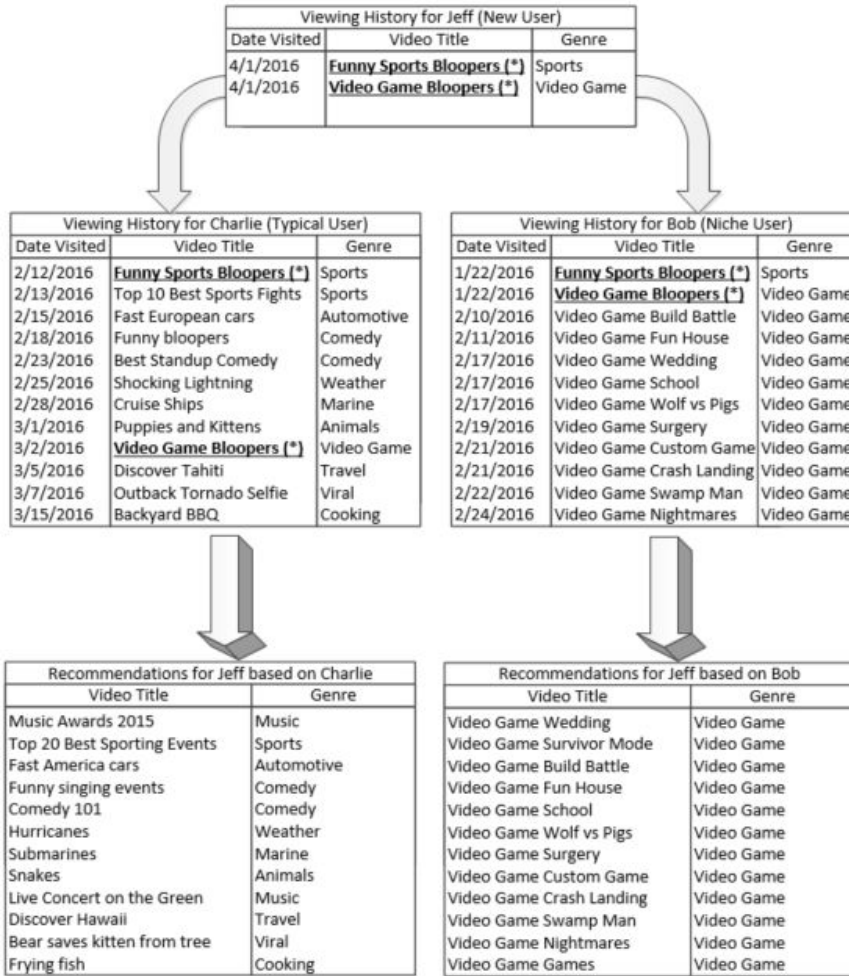
Videos in the first category primarily drive their views from frequently returning users, such as those that watch episodic content, vlogs, and niche content. These frequent viewers tend to have a constant and predictable watch pattern because of their eagerness to watch the latest scheduled release. In addition, frequent viewers are willing to spend a considerable amount of time watching the content.

Videos in the second category primarily drive their views from infrequent viewers, such as those that watch viral videos, political videos, and news videos. In

this context, “infrequent” may refer to lack of a tendency to watch videos from the same genre, rather than frequency of viewing; some “infrequent” viewers may constantly watch a large number of videos, albeit across random genres without following any particular topic. These infrequent viewers also may not sign-in under an assigned social media username, which makes their watch history unavailable to the social media site’s video recommendation system. Consequently, the social media network cannot use the co-viewership algorithm to identify video content that is relevant to the user’s interests. Instead, the social media network may switch to an alternate recommendation algorithm, such as one that is based on the overall watch time for a video - similar to attaching an average characteristic to the anonymous user. For example, a first social media user may watch a 30-second video on swimming, a 60-second video on soccer, and a four-hour video on golf. Based on the significant watch time of the golf video as compared to the watch time of the other two videos, the social media network will recommend the golf video (or videos about golf) to other social media users who also watch sports videos.

### **Over-Recommending Niche Videos to Infrequent Viewers**

The inability to identify a frequent user who views niche content from an infrequent user who also views the same or similar niche content can have a significant impact on the quality of the video recommendation algorithm. For example, certain niche content contributes a disproportionate amount of watch time to the recommendation algorithm because the fans of this content tend to watch for long periods of time and return frequently. As a result, the social media network’s recommendation system may begin to over-recommend these videos to the other infrequent users who may not be equally as interested in the niche content, but made the innocent mistake of watching some of the niche content.



The diagram above shows two possible video recommendation lists for Jeff, a “new” user of the system. The bottom of the left-most branch of the diagram shows the recommendation list for Jeff based on Charlie’s viewing history, where Charlie is a “typical” user of the system. The bottom right-most branch of the diagram shows an alternate recommendation list for Jeff, but based on Bob’s viewing history, where Bob is a “niche” user of the system.

As shown, Charlie’s viewing history from 2/12/2016 to 3/15/2016 includes a variety of different viewing content. For example, Charlie watches content from various genres such as those that fall under Sports, Automotive, Comedy, Weather, Marine, Animals, Video Game, Travel, Viral Videos, and Cooking. Charlie’s viewing

history is typical for most users visiting the system in that he does not repeatedly watch content from a select few genres.

Conversely, Bob's viewing history for a similar duration of time includes only videos that fall under the single genre of Video Games. Bob's viewing behavior is typical for a niche viewer in that he spends a substantial amount of time viewing content of a similar type.

When the system identifies content for a new user (such as Jeff) based on a typical user (such as Charlie), the quality of the recommendation is generally high. For example, Jeff's viewing history includes "Funny Sports Bloopers," a video in the Sports genre, and "Video Game Bloopers," a video in the Video Game genre. Since Charlie has also watched these two videos, the recommendation system determines that Jeff and Charlie have similar interests. Accordingly, the system may recommend for Jeff to watch the other videos on Charlie's viewing history, and in doing so, Jeff will have a recommendation list that contains video content from a normal balance of genres that are likely to be within Jeff's interests.

However, the video recommendations for a new user of the system may be skewed by a heavily engaged niche user (such as Bob). For example, the system will recognize that both Jeff and Bob have similar interests because both watched "Funny Sports Bloopers" and "Video Game Bloopers." As such, the system may recommend for Jeff to watch the other videos on Bob's viewing history. Yet, Bob is a niche-user with an atypical interest in video game content. Thus, the system will over-populate Jeff's recommendation list with content from a single genre. Consequently, assuming Jeff is not also a niche user (which may be statistically unlikely), the system will likely fail to compile a list that accurately reflects Jeff's actual interests.

To address the above problem of over-recommendation, specific restraints have been added to the system to exclude known niche content that interferes with the quality of the recommendation algorithm. However, this is not a viable solution

because the restraints require constant manual updating in order to keep up with the high volume of new niche content continuously being added to the system by its users..

### **Scaling Contributions Based on User Type and Content Type**

A social media recommendation system may adjust the recommendation scores assigned to videos by scaling the watch time contribution from a user differently based on the type of user, the type of content viewed by the user, or both. This approach, for example, can minimize video recommendations skewed by a heavily engaged niche user. That is, the system will recommend videos that are more relevant to each users' interests. More relevant recommendations are critical for reducing the number of search queries that users perform to find videos that the user wants to watch. This has the benefit of not only reducing the amount of network traffic on the site, but also improving the battery life consumed in the search process.

In one implementation, referred to herein as a "simple" approach, the system takes into account the total frequency of visits of the user in order to segment the watch time into two categories: watch time from frequent visitors and watch time from infrequent visitors. For example, each time the user visits the site, the system will log the user's device identifier. While this identifier does not contain the user's personal information, it may be used by the system to determine if the user of the device is a frequent visitor of site.

By categorizing the user as either frequent or infrequent, the system is able to scale each user's watch time contribution differently. For example, as shown in the above diagram, Bob viewed niche content from 1/22/2016 to 2/24/2016. Based on Bob's viewing behavior during this time period alone, the system may consider Bob as a frequent user. Accordingly, Bob's viewing behavior of viewing only niche

content may be appropriately scaled such to avoid over-recommending video game content to a new user, such as Jeff.

Nevertheless, the approach of scaling watch time differently based on a single factor, such as the type of user, may not always avoid the over-recommendation problem described above. Assuming Bob never revisits the system after 2/24/2016, the system may decide to treat Bob as an infrequent user. As a result, Bob's obsessive video game watching behavior will impact the recommendation list for Jeff.

For example, as shown above, Jeff visited the site on 4/1/2016, which is after the last time Bob viewed a video about video games. The system may decide that as of 4/1/2016 (when Jeff visits) that Bob is not likely to return to the site because he has been gone since 2/24/2016. Accordingly, Bob's contribution as an infrequent user would not be appropriately scaled, even though his obsessive viewing of video games is atypical for most users of the system. Consequently, Jeff's recommendation list would be over-populated with video game content.

Conversely, the system could possibly consider Bob as a frequent user based on his obsessive viewing behavior during his four-week visit to the site. Or, Bob could have visited the site between 1/22/2016 and 2/24/2016 when Jeff was actively using the site as a new user. In both instances, the system may consider Bob as a frequent user and negate his watch time contribution when populating Jeff's recommendation list.

In an alternate implementation, referred to herein as a "sophisticated" approach, the recommendation system segments the watch time into several categories, which reflect the frequency with which users watch content from the same genre as the video. This approach has the advantage that it can distinguish between (1) watch time generated by niche users within their niche and (2) watch time by the same users for content that they do not visit frequently. As a result, the system may further distinguish between the watch time of a niche video by a fanatic of the niche

video and the watch time on a viral video from the same user. The advantage of this approach is that it enables the system to skew the recommendation scores such that the watch time from the viral video is applied with a larger multiplier than the watch time from the niche video.

For example, as discussed above, the system characterized Bob as a fanatic of niche content because his viewing history included content from only a single genre – video games. As such, the system scales Bob’s watch time results when calculating Jeff’s recommendation list so to avoid over-recommending video games.

On the other hand, if Bob’s viewing history also included a single viral video, then the co-viewership model suggests that Jeff would also be interested in that viral video because both Jeff and Bob watched “Funny Sports Bloopers” and “Video Game Bloopers.” However, by scaling based on user type alone, the system will fail to recommend the viral video to Jeff. Thus, scaling based on both user type and content type will improve the quality of the recommendation system for users such as Jeff.

### **Example of operations**

In one implementation, the system has a number of phases that can run in parallel. The system first identifies the type of user as either frequent or infrequent. If the sophisticated approach is used, then the system will also identify the type of content viewed by the user as either niche or non-niche.

#### a) Identify Type of Content

The system may identify at least one genre of content the video belongs to. In one implementation, this may be predetermined by an algorithm or may be determined based on flags defined by the content creator. For example, the content creator of a twitch video generally specifies the title of the game being played.

#### b) Identify Type of User



The system may identify at least one genre of content the viewer frequently watches. In one implementation, this may be predetermined by an algorithm that counts the total number of video views or watch time of videos for each genre over a given period of time. For example, the system may determine if the user has 5 minutes of watch time from videos from the same channel for at least 3 of the last 30 days.

Users might be categorized into various frequencies. For example, a fanatical user might have watched at least 10 videos in the last 30 days, but a regular user might have only watched at least 5 videos in the last 30 days. In some applications, the frequency of use might be described as a scalar value such as “at least X days with 5 minutes of watch time over the last 30 days.”

c) Identify Intersection of Genres

The system may find the intersection of genres between the viewer and the video. In one implementation, only the highest valued genre will be considered. In other implementations, some applications might add the value of the genres together. However, the latter approach may introduce a few loopholes and other undesired outcomes.

d) Generate scalar value for ranking

The system may produce a scalar value from the combination of the watch time and the frequency of views by the user for the specific genre.

In some implementations, the scalar value may be used to rank content or be used as a factor in a more sophisticated ranking algorithm. In other implementations, the system may divide the scalar value by the sum of the scalar values for all video views to identify a percentage of total revenue to attribute to the video view.

Instead of genre, the system may be applied to particular channel or other more granular classifications of a video such as playlists. In some implementations, the system may use videos watched by a particular user cluster as the splitting criteria. These viewer cluster-defined buckets are useful because they may describe a wide variety of genres which humans are unlikely to be able to describe or even count.

Accordingly, this disclosure provides details regarding implementations of systems for a video recommendation algorithm to scale the watch time contributions from frequent viewers of niche content.