**Tianen’s evaluation metric suggestion on Bilibili’s recommendation system**

**Introduction**

Bilibili is a popular Chinese video-sharing (other forms of content also exist, but video remains the largest portion by far) platform, its similar at core with other video-sharing platform like YouTube but it offers some distinctive features:

* Bullet Comments (弹幕): Comments made by user A at a specific time point during a video would show up on user B’s screen when he/she reaches the same time point in the same video.
* Coins (硬币): A virtual currency designed to be used as a way to evaluate and promote excellent content, not for trading.
* Moments (动态): Postings that allow users to tweet&retweet, upload videos/articles, and for content creators who also do streaming on the platform they can also notify their followers that they are now live through moments postings.

These are only some major example features that Bilibili has, and together they made Bilibili not only the largest Chinese video sharing website, but also the biggest Chinese online ACG (anime, comic, games) community and one of the most frequently used social platform.

**Recommendation System Considerations**

Although Bilibili never official disclose any insights on their approach to recommending videos, video creators have came up with formulas that they concluded based on their experience with the platform:

Recommendation Weight = Base (Account) Weight + (A \* Like Ratio + B \* Coin Rate + C \* Collection Rate + D \* Share Ratio + E \* Bullet Screen + F \* Comment) - G \* (1 - Completion Rate) - H \* (1 - Click-through Rate)

A - H here represent possible weights that Bilibili assigns to each attribute, for rates and ratios, they stand for the percentage of people who decided to take certain actions after watching a video, for instance, like ratio stands for the percentage of people who liked a certain video after watching it, and coin rate means the percentage of people who decided to give coins to a certain video after watching it, and so on.

**Primary Concern**

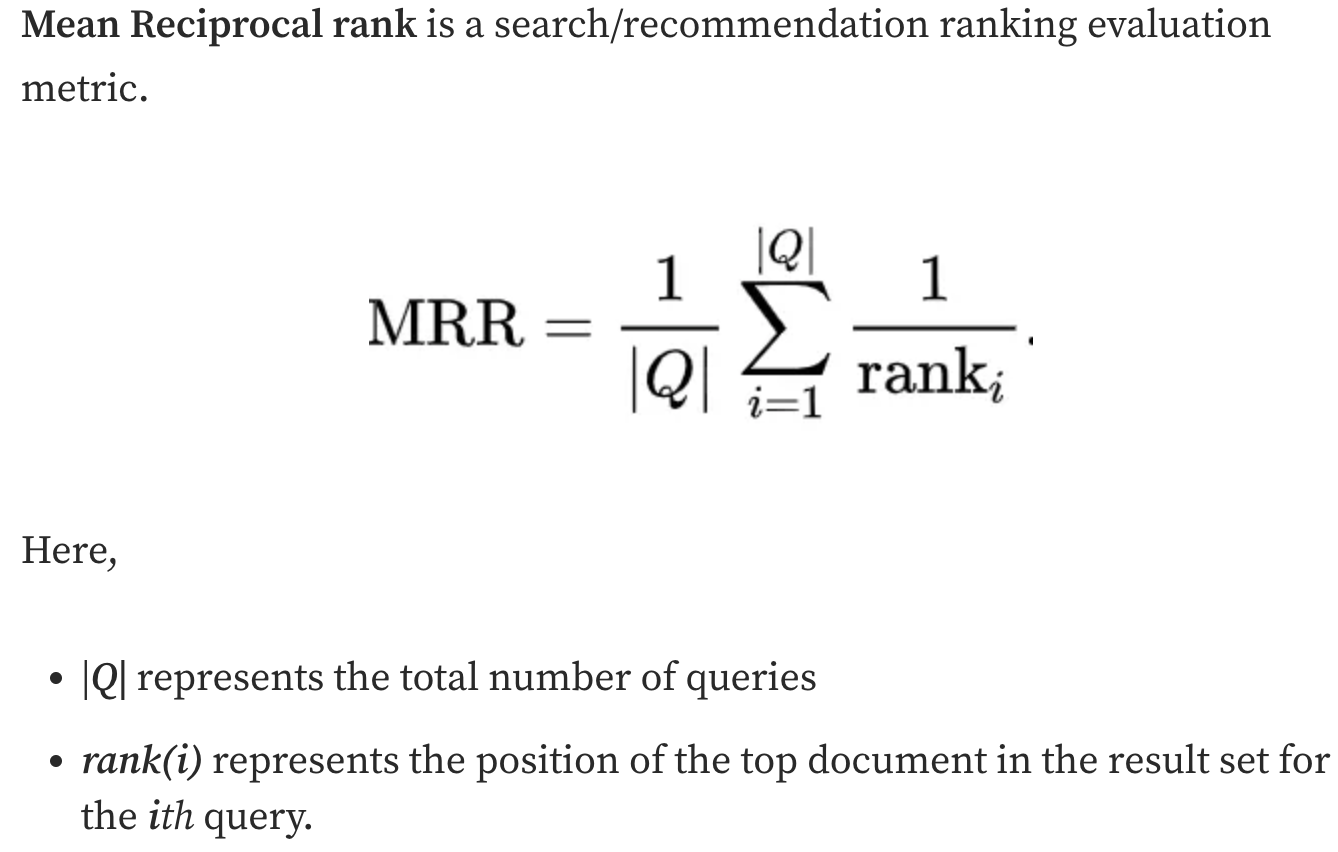
Since bilibili is a video sharing platform at core, when designing a recommendation system the main objective would be minimizing false positive and maximizing true positive, as all the videos that we are recommending are ‘positive’, and among those ‘positive’s we want every one of them to be accepted by our user. And this means the primary focus of evaluation on Bilibili’s recommendation system would be improving precision.

**Suggested Metrics**

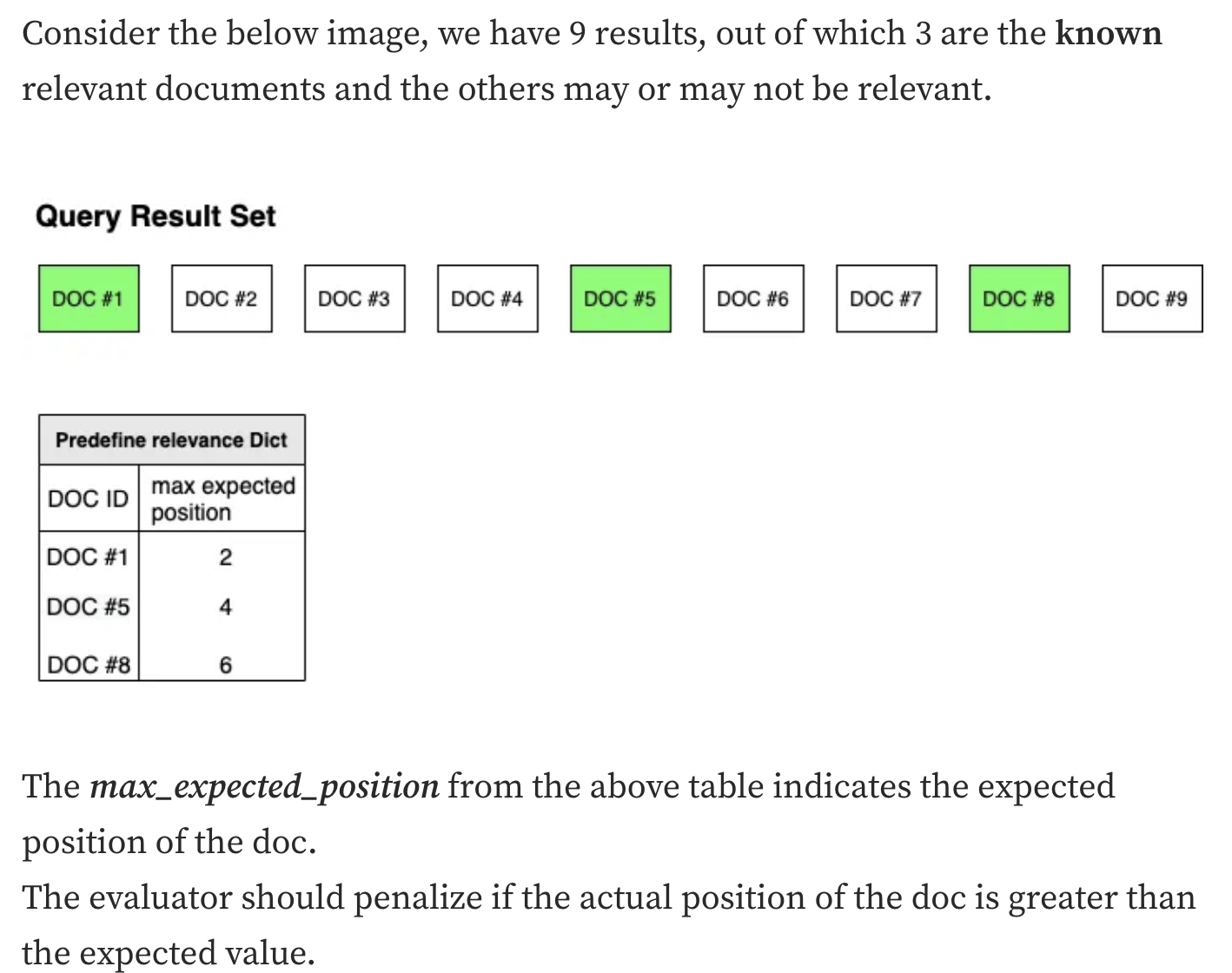
Specificity: In addition to the primary concern of improving precision mentioned above, Bilibili could also make measures to lower false negatives, since a false negative is an actual positive and if Bilibili predicted it correctly it would add to the total number of true positives, improving Bilibili’s recommendation system. And to keep track of this, Bilibili could focus on maximizing specificity, which measures true negative rate by calculating: True Negatives / (True Negatives + False Positives).

Extended Reciprocal Rank: Another metric that I found very appropriate for enhancing Bilibili’s evaluation is ‘Extended Reciprocal Rank’ shared by software engineer Mr. Bikas Katawal on ‘medium.com’ (link provided in references below). In his post on ‘medium.com’, he explained how he extended the concept of mean reciprocal rank to n known relevant documents instead of just focusing on the top one. He also provided graphs and formulas to help better understand the method:

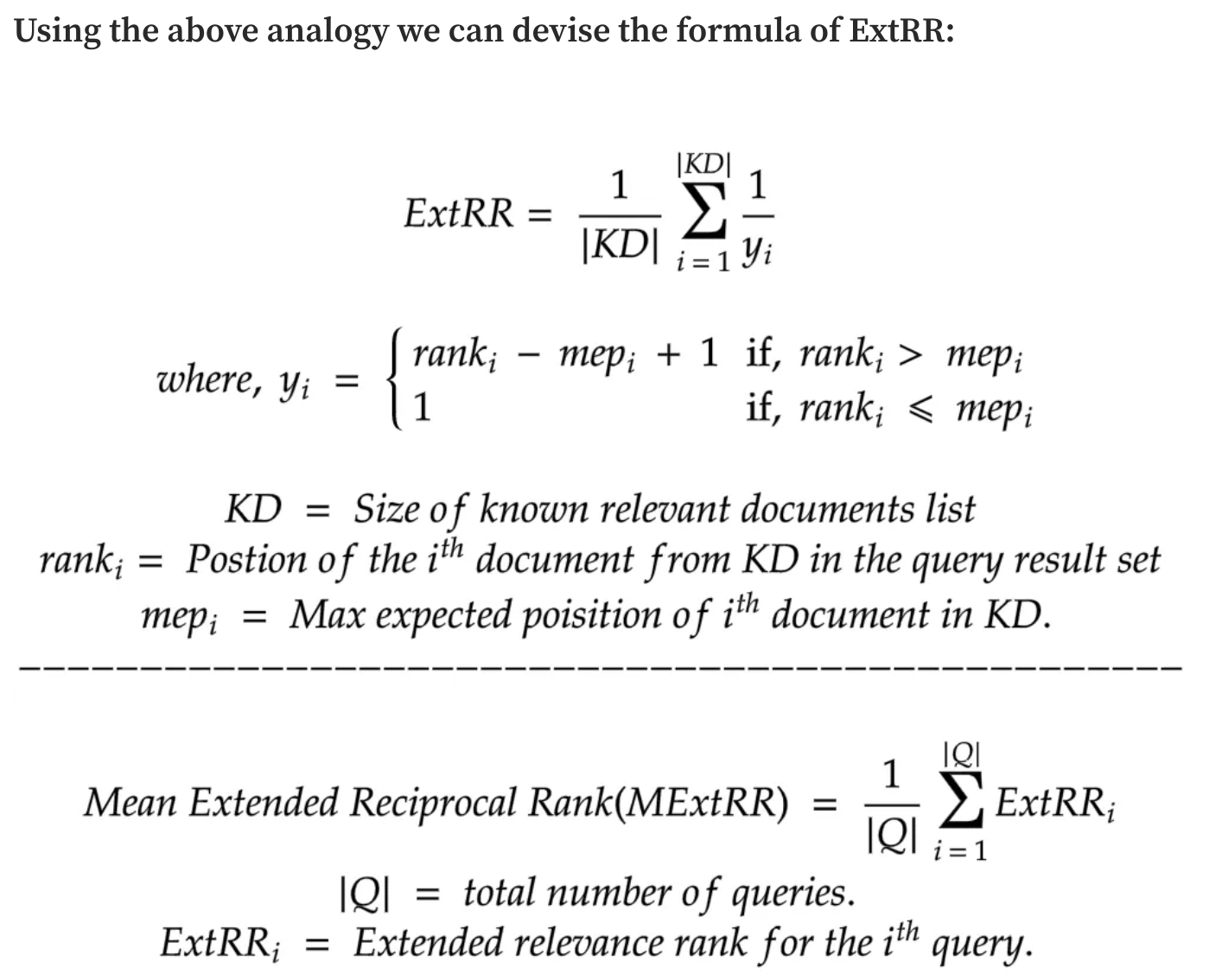
He first explained MRR



Then provided an analogy to help audience ease into the concept

****

Then he gave the formula to the approach:

****

**Suggestions from Personal Experience**

Since I’ve been using the platform for several years, I have some suggestions for them coming from personal experience as well:

When considering the impact of users adding a video to their collections, it may be valuable to incorporate the number of collections in which a user places a particular video. As an end user myself, I often myself adding exceptional videos to multiple collections, while adding some others that are not as good to only one collection. I personally think this is plausible since videos that excel often offer insights across various subjects or fields, and taking this into consideration could really separate the exceptional videos from generally good ones, making Bilibili’s recommendation system more refined.

I have also observed that most users, including myself, typically accumulate coins through consecutive daily log-ins, receiving one coin for every two days. However, the minimum number of coins that can be contributed to a video is one, with two coins often encouraged as the maximum single-video contribution to boost its visibility in Bilibili's recommendation system. This would ultimately result in users running out of coins when encountering a high-quality video they wish to support, potentially forgetting to return later to contribute coins (happened to me many times, I would suddenly remember to contribute coins to a video weeks later). On the other hand, users with an abundance of coins may contribute to less outstanding videos simply due to their surplus. Thus, I suggest that Bilibili dynamically monitors the activity of coin contributions, tracking not only the number of coins awarded to a video, but also the user's coin balance at the time of a contribution. This data could be used to optimize the recommendation system and better reflect users' genuine appreciation for content.

**References**

User concluded general approach to Bilibili’s recommendation system: https://www.bilibili.com/read/cv10011552

Introduction to Extended Reciprocal Rank: https://towardsdatascience.com/extended-reciprocal-rank-ranking-evaluation-metric-5929573c778a