Nicholas Belev

Professor Santiago Gil Begue

Chatbots and Recommendation Engines

8 February 2025

Assignment 3 - Performance of Various Recommender Systems

Accuracy:

Proposed Performance Ranking (1 - Best; 4 - Worst) :

1. Collaborative Filtering
2. Content Based
3. Popular
4. Random

Justification:

Allow us to first define the accuracy of a recommender system as the level of alignment or similarity between the content and ranking that the system recommends (and the confidence it has that it is to the user's taste) and what the user enjoys seeing (how they would actually rank the recommendations). This can be quantified using precision, recall, or Normalized Discounted Cumulative Gain (NDCG).

Abiding by this definition, we come to the most evident ranking choice: it is expected that a Random recommender will undoubtedly perform the worst. This is because a recommender that shows content according to a random variable will undoubtedly give exceedingly strong weight to obscure content that would only be appropriate for a select few users. Whereas, a random system has no "knowledge" to discern if the current user is an appropriate candidate. A uniform random recommender—with essentially zero regard for user preferences influencing its decision—is one of the worst possible recommenders. Even if the random variable is a distribution proportional to the popularity of items, it will still (with above zero probability) recommend fully irrelevant items to users on occasion.

Next, a Popular recommender is informed by what the system knows is trending in terms of quantity of viewers, frequency metrics, amount of reviews—a metric by which the recommender knows an item is generally liked and heard-of by the majority of the population. Consequently, by only recommending what is popular, the system increases its chances that it is showing an item that will also be liked by a specific user, because "most people do". Nonetheless, this is still a naive method, with a possibility that the user does not like this popular item, and with no element of personalization to provide other more informed, user-specific suggestions.

In terms of user specificity, this is an element that Content-Based recommendation systems account for. This RS has paid attention to the specific user's propensity to interact with items based on a typology that they have been tagged with or grouped by. Consequently, it has a better idea of the "types" of content that are of preference to the user. Utilizing this proportional breakdown of the user's taste, the content-based recommender can show these types of items, to add some variety and customization to niche categories beyond those that the popular content falls under. Its only flaw is that it still needs some underlying distribution with which to select and rank items based on which categor(y)/(ies) they exhibit, so will not always be accurate (with its choices of what a user would not be interested in, and thus what is not shown to them).

Lastly, Collaborative Filtering recommendation systems improve further upon Content-Based recommender systems because they have the ability to gauge profile interests for individual users and further, to analyze the similarity between users, so that in the event that User A shows an observable interest in Item X that perhaps the system wouldn't have naturally recommended to User B (who has previously demonstrated preference similarity to User A), the recommender would know that User B would also have a high probability of liking this item X. This allows for a level of personalization and therefore accuracy (and confidence in this accuracy), significantly higher than the other RS's.

Coverage:

Proposed Performance Ranking (1 - Best; 4 - Worst) :

1. Random
2. Collaborative Filtering
3. Content Based
4. Popular

Justification:

Let us define coverage as the proportion of available content on the platform that the recommender will show (or make available) to the population using the platform. It ensures that users have a wider selection and freedom of what content they can reach, but must be balanced with accuracy.

It is for this reason that Random RS has the lead, since Random selects items without preference bias, it gives significantly more exposure to low-ranked or rarely chosen items—content that would otherwise be ignored by more targeted recommendation engines. Conversely, even with a distribution proportional to the popularity of items used to recommend them, the standard deviation of this is still higher (more varied random recommendations, even rarely-interacted-with items) than that of an engine that has better zeroed-in on the profile and tastes of a user. There is thus no (or less) bias to showing some items over others.

Collaborative Filtering RS follows suite, it can figure out a user based on their interactions and even find something new and relatively unpopular to show them, if similar profiles to them have expressed interest in such an item, making it able to confidently suggest a wider breadth of items to all users. Of course, collaborative filtering is subject to the "cold start" problem, since it needs to have an understanding of all the profiles in the network / platform, as well as gauging a new user before it can show them niche (perhaps unexplored) content, while performing well in the accuracy metric (its primary goal). Nonetheless, collaborative filtering still has a tendency to recommend popular items (an impact of the long tail phenomenon), and so its diversification is limited.

Content-Based RS, though they have a similar principle to Collaborative Filtering, lack the crucial element that helps them offer more diverse options: recommendations based on similarity to other users. As a result, content-based recommenders rely entirely on the user to make searches and interact on their own for the system to know what kinds of items to show them, and even then, it will have a bias towards popular and similar items in those preferred categories (they are less risky picks for the RS), since it wants to maximize the probability that this user will like an item from within the category. This means that the user is less likely to be shown a niche item than if a similar user to them interacts with that same item, and so coverage will not be able to be increased as greatly with a content-based system.

To put it briefly, a Popular RS, recommends only what is popular and what is trending (the narrow head of the long tail phenomenon curve), this means users are being shown only the elite minority of content on the platform and the 'long tail' remains uncovered by this recommendation system. It accounts for all users as a collective audience by recommending the most widely engaged content. It determines this content solely on interaction metadata and aggregate engagement metrics, which limits its coverage to only frequently consumed content; clearly it has the lowest coverage.

Personalization:

Proposed Performance Ranking (1 - Best; 4 - Worst) :

1. Collaborative Filtering
2. Content Based
3. Random
4. Popular

Justification:

To begin, personalization ought to be defined as how different recommendations are between different users. In other words, how unique is any particular user's "for you" page? A maximum score in personalization means that, for any two users, their lists of recommendations are almost entirely different, whereas a low score indicates extremely similar recommendations. Note how this interacts with coverage and accuracy: if two users have highly similar tastes, then by our definition here, their recommendation lists would end up looking quite similar and thus be considered "less personalized", even if it is accurate. Of secondary note is the increased difficulty of obtaining personalization when the platform has a high number of users.

It follows that a Random RS emerges as the best system for personalization precisely because it has no structured logic for assigning items beyond random chance. By default, each user is more likely to receive a unique selection of items—there is minimal overlap among users' recommended lists because there is nothing steering the system toward recommending the same subset of items to everyone. Of course, this randomness implies almost no accuracy (an inverse relationship noted earlier, as with coverage), but it excels in ensuring that any two users see highly distinct recommendations.

Collaborative Filtering ranks below Random. While it does indeed look for patterns of similarity in user behavior, it does so in a manner that still allows for some level of uniqueness (while maintaining accuracy): two users who share certain tastes might overlap in a subset of recommended items, but they will likely still receive individualized suggestions based on nuanced differences in their interaction histories. This yields more personalization compared to systems that do not factor in user preference at all (or that factor it superficially), yet it typically cannot outstrip a pure Random RS on the "uniqueness" dimension alone, because randomization introduces the greatest separation in recommended lists from one user to another.

Regarding Content-Based RS, it also offers some degree of personalization but falls short in comparison to the above. By focusing exclusively on what each individual user has interacted with, it can produce recommendations that match a user's established tastes, but if two users share remarkably similar categorical interests via interaction history, their "for you" pages will end up looking very much alike.

Additionally, since it ignores the tastes of other (possibly unique-taste) users, it misses an extra chance at uncovering interesting outliers—items that might yield further uniqueness in one's recommendation list.

Popular RS essentially delivers identical recommendations to everyone by focusing on whichever items are trending or have the highest interaction counts. Because popularity metrics drive it, it fails entirely at personalization under this definition—every user is fed the same top hits or most-viewed content. As a result, there is little to no variation in recommendations across users, leading to the lowest personalization score..

Diversity:

Proposed Performance Ranking (1 - Best; 4 - Worst) :

1. Random
2. Content Based
3. Collaborative Filtering
4. Popular

Justification:

To begin, we define diversity as how dissimilar the recommended items are in terms of their categories or metadata. In simpler terms, if a user just watched a James Bond movie, a system with low diversity would only suggest other James Bond films or closely related spy thrillers, whereas a system with improved diversity might also include historical dramas, documentaries, etc. Typically, like with coverage (on a population scale), an extreme level of diversity (on an individual scale) harms accuracy by deviating from user preference.

It follows that a Random RS is the highest diversity system exactly because it places no constraints on the types of items it selects. Each user sees a wide assortment of categories picked at random, giving more chances for rare, obscure items to appear. This unfiltered variety maximizes the probability of encountering a high spread of categories or content types for a single user and translates to the highest diversity ranking.

Content-Based RS ranks second. While it focuses on the user's demonstrated interests, it can still cover multiple facets of those interests if the user has interacted with a range of different categories. For instance, if a user has shown an affinity for a number of general categories, the system can recommend titles across those broader categories. However, it lacks external input from other users' diverse tastes, which slightly constrains its recommendations. In other words, it only ventures as far as a user's existing profile allows, limiting the potential for individual variety.

Collaborative Filtering ranks worse than Content-Based. In principle, its ability to draw similarity from users can introduce some diversity—especially if one's "neighbors" in the user-item matrix have sampled a varied assortment of items. Yet, collaborative filtering has a tendency to zone in on popular or overlapping interests within these user clusters, creating pockets of homogeneous recommendations. This clustering effect can inhibit the system from serving up the most diverse possible range of categories compared to the purely user-centric and random approaches.

Lastly, the Popular RS ranks as the least diverse because it prioritizes only those items that have garnered the most interactions overall. Every user sees essentially the same top trending content—from the same small set of mainstream genres or categories. Even if there are countless niche genres on the platform, the popular recommender doesn’t show them, so for the same reason as coverage, a Popular recommender performs poorest for diversity.