'Fitness that Fits': A prototype model for Workout Video Recommendation

Ercan Ezin
Computer Science Department
University of Bristol
Bristol, United Kingdom
ercan.ezin@bristol.ac.uk

Eunchong Kim
Computer Science Department
University of Bristol
Bristol, United Kingdom
ek17843@my.bristol.ac.uk

Iván Palomares
Computer Science Department
University of Bristol
Bristol, United Kingdom
i.palomares@bristol.ac.uk

ABSTRACT

Personalization services enable Internet users to benefit from tailored recommended content at their finger tips. Our interest in this contribution lies in video recommendation within the fitness domain to support an active lifestyle. We present 'Fitness that Fits', a preliminary platform for workout video recommendation, which benefits from the Youtube-8M labeled dataset and its rich variety of categorized video labels, thereby enabling fitness workout video recommendations predicated on the users' preferences and their recent viewing behavior. The proposed model also incorporates an approach to produce diversified recommendations and foster the practice of distinct fitness activities based on like-minded users' information. An initial experimental study shows the trade-offs of the hybrid approach considered.

KEYWORDS

Personalized Wellbeing; Preference Modeling; Diversity

ACM Reference Format:
Ercan Ezin, Eunchong Kim, and Iván Palomares. 2018. 'Fitness that Fits': A prototype model for Workout Video Recommendation. In *Proceedings of the Third International Workshop on Health Recommender Systems co-located with Twelfth ACM Conference on Recommender Systems (HealthRecSys'18), Vancouver, BC, Canada, October 6, 2018* , 6 pages.

1 INTRODUCTION

The Internet and its associated technologies have become an indispensable tool to search products, services or frequently access information needed in our daily lives, e.g. booking a hotel, purchasing a new device or consulting the weather forecast. We are presently reported to spend an average 6 hours per day connected to the Internet. Amid this phenomenon, there is an increasing interest in seeking aid in the Internet to embrace healthier lifestyles, e.g. through the search and sharing of information related to fitness exercises and wellness practices, or via smartphone apps [10]. For instance, the rate of Google searches based on keywords such as "personal trainer", "crossfit", "hiit", "core" or "pilates" has dramatically increased in the last decade [2].

Although gyms and leisure centers are a common choice for users who desire to adopt or maintain an active lifestyle, they are not always within the reach of every person, e.g. owing to financial limitations, busy schedules, frequent traveling, etc. Taking advantage of the growing demand for online resources to promote exercising, online workout videos have proliferated in recent years as an alternative means to keep users active from the comfort of home or beyond, with a number of advantageous characteristics:

- They are convenient, providing 24/7 access to a wealth of fitness resources from anywhere with an Internet connection.
- They do not require commitment to work out at an externally imposed day or time.
- With a careful search and use of the resources available, they provide a wealth of workouts from a diversity of instructors.
- They are cost-effective and can be undertaken in a more individual and private space.

Whether it is bodyweight workouts, aerobic exercises, performance hacks or skill gaining tutorials, Youtube provides millions of users with access to a wealth of video resources to support them in practicing their preferred workouts anywhere and anytime. Despite the potential benefits to Youtube users of receiving personalized recommendations from the platform as a whole [5, 6] and subscribe to specialized Youtube channels, the popular Internet platform does not provide an exhaustive categorization of workout videos into different types of exercise and sport. Whilst this is not a critical issue for most users interested in a sheer variety of videos and themes, nowadays there is a growing niche of users specifically interested in accessing fitness videos to a considerable extent. These users would benefit from a bespoke service for fitness workout video recommendation, that exploits categorized (labeled) video data describing the types of activities shown in such workout videos. Some studies focus on recommending video resources related to healthcare [11] and, more specifically, fitness [3]. Notwithstanding, state-of-the-art works on fitness video recommendations mostly rely on small video data sets that have been carefully selected by a domain expert. This causes any model implementation to be hardly extrapolatable into a large-scale setting, making them poorly generalizable [11].

The Youtube-8M dataset is a clear example of large-scale dataset containing comprehensive information about millions of videos uploaded to Youtube. Despite having been primarily investigated for tasks such as the classification and further understanding of video data [4], it has barely been used for recommendation processes to date. As a result of classification and supervised machine learning processes on data originating from Youtube videos, Youtube-8M incorporates labels associated to the videos, thereby describing the topic(s) to which they belong, including a number of fitness activity types: this amount of labeled video data has an untangled potential to investigate and enhance existing recommendation approaches on large volumes of video related to specific domains such as fitness.

HealthRecSys'18, October 6, 2018, Vancouver, BC, Canada

To overcome the challenges outlined above, this contribution presents 'Fitness that Fits', a prototype platform for workout video recommendation, which relies on Youtube-8M video data describing fitness activities. Our main contribution is a recommendation model that extends principles from content-based and collaborative filtering by introducing mechanisms to provide end users with *meaningful* and *diverse* workout video recommendations.

2 RELATED WORK

Recommender systems applied to health-related domains are still relatively scarce [8], particularly in the area of fitness and general wellbeing. This section reviews some representative recent works on recommender approaches in these domains. Following this, we briefly describe specific models targeting the fitness domain, along with similarly purposed models for video recommendations.

Ceron et al. [3] presented *CoCare*, a mobile-based recommender system aimed at supporting health promotion and preventing diseases, by recommending physical activity videos based on users' profiles and their context. Their approach relies in a decision tree learning algorithm and a Case-Based Reasoning (CBR) module with a twofold purpose: (i) classifying and tagging videos predicated on their textual description, and (ii) calculating similarity values between user profiles and video categories. The system presents the limitation of not exploiting the vast and hugely accessed videos from Youtube, which as outlined by the authors, would required a proper categorizing and profiling process to make the recommendation process suitable to the specific domain. Instead, it relies on a small assortment of videos selected by expert users. This is indeed a limitation present in other approaches to health-related video recommendation [3, 11].

To assist health professionals, patients and caregivers in the process of finding relevant information amid a plethora of it, Sánchez-Bocanegra et al. [11] built *HealthRecSys*: a content-based recommender system aimed at providing personalized links to educational content that supplements online health videos. Extracted text from metadata in relevant Youtube videos is analyzed to generate Medline Plus[1] links. Expert information from healthcare professionals is required to gather a relevant dataset of videos, such that multiple experts rate the quality and relevance of recommended links for given videos, and only those videos and links showing consensus among experts are selected for the experimental study of the system feasibility. This human effort is motivated by the need for mitigating potential risks for health consumers.

Within the particular scope of the fitness domain, two studies on recommendations for running were presented by Berndsen et al. [1]. In, [1] the authors investigate the performance and progression differences between casual and elite runners, examining the feasibility of a recommender system for runners. The importance of providing runners with explainable recommendations and using resources to keep them motivated, are particularly highlighted. A different approach on fitness activity recommendation was adopted by Dharia et al. in [7] , with a focus on social recommendations for personalized assistance in performing fitness activities: their approach integrates mobile or wearable-based activity data, preferences, goals and contextual information to produce socially enhanced fitness recommendations.

3 DATA AND SYSTEM OVERVIEW

This section describes the currently used data in 'Fitness that Fits' and briefly overviews its Web platform being developed.

YouTube-8M is a large-scale labeled video dataset which, as of June 2018, consists of over 6 million of YouTube video instances (which add up to 350,000 hours of video), namely video IDs with high-quality annotations generated by machine learning techniques, describing a highly diverse vocabulary of over 3.8K different entities (labels). Each video in the dataset contains an approximate average of three labels. These videos have been sampled uniformly with the aim of preserving the highly heterogeneous distribution of popular Youtube contents, whilst ensuring the dataset quality and stability by enforcing a series of restrictions. The dataset also incorporates precomputed audio-visual features from billions of frames and audio segments, which facilitates an efficient training of baseline models without the need for sophisticated computer settings, and (ii) enables a deep exploration of complex audio-visual models that otherwise would be impractical to train.

The use of Youtube-8M has been illustrated in recent research efforts including large-scale video classification [12], multi-label classification [14], extraction of the hierarchical structure associated to Web video groups [9], etc. To the best of our knowledge, however, Youtube-8M labeled video data has been barely investigated within the scope of recommender systems research, despite its potential advantages for fitness workout video recommendation:

- Its comprehensive topic information (labels) resulting from previous classification efforts by the Youtube-8M community, constitutes an added value for highly personalized video recommendation.
- Topic information can be further exploited to promote diversity in recommendations.
- It provides a large and rapidly increasing volume of videos from the Youtube platform.

We built an initial dataset by filtering the original Youtube-8M labeled video dataset [2] predicated on the following filtering criteria:

1. *Highly-viewed*: Only videos with a minimum of 50K views in Youtube are selected for the scope of the preliminary research presented in this contribution. Whilst important, tackling the popular cold-start problem associated to newly added content lies outside the scope of the present study.
2. *Fitness-related*: Videos having machine-generated annotations of 'Beauty and Fitness' narrowed down to 16 labels, associated with popular and highly-viewed types of fitness activities in accordance with criterion (1).

The resulting video dataset comprises 1.7K fitness workout videos with over 1K user views each, which supposes an elevated number of videos in contrast to other related works [3, 11].

Figure 1 provides an overview of the user interface for the 'Fitness that Fits' Web platform. Besides the interface for exploring videos or viewing the recommendation list, the user can establish and update anytime a profile by selecting at least two labels as her

[1] Medline Plus website: https://medlineplus.gov/

[2] https://research.google.com/youtube8m/

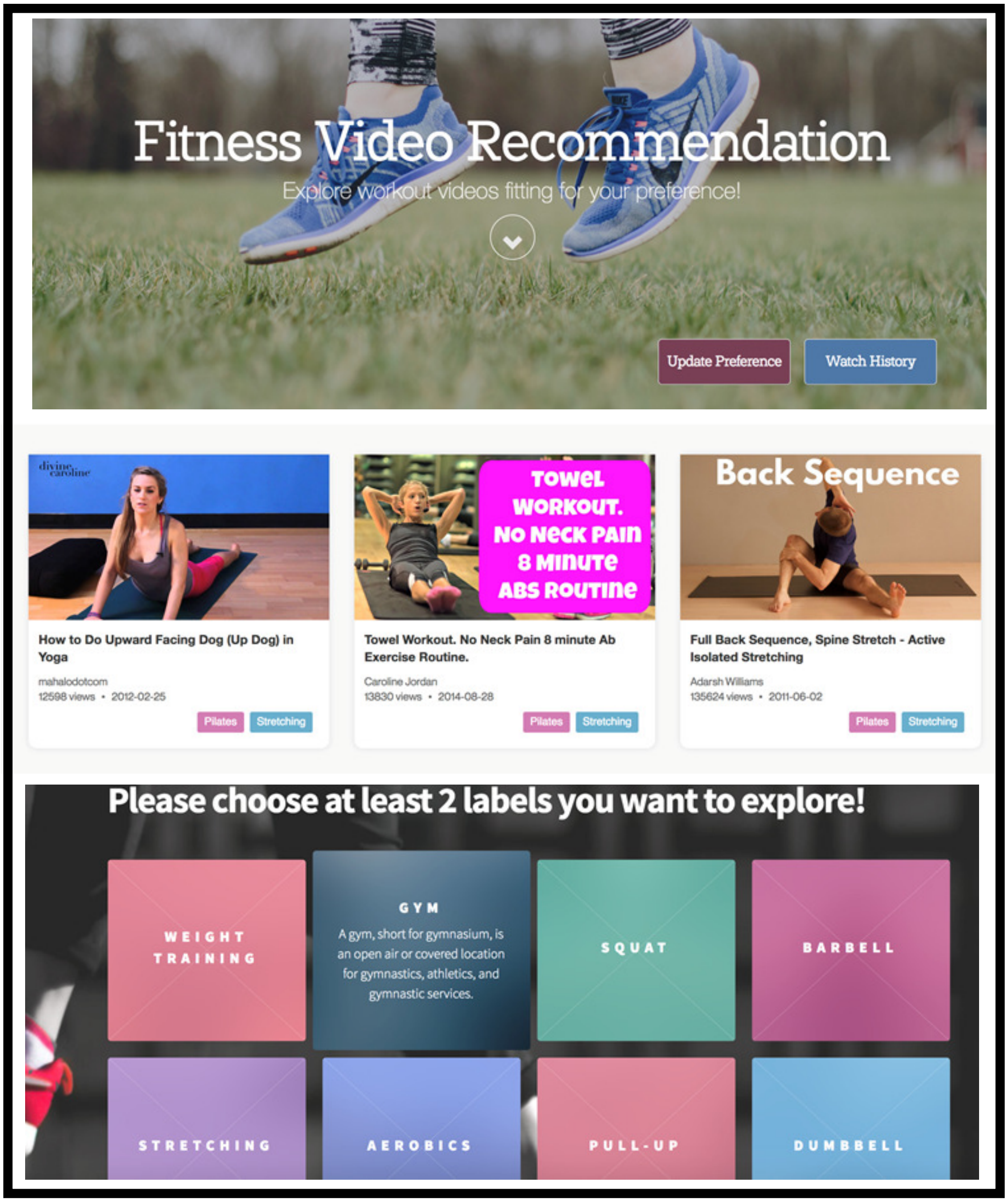

Figure 1: Appearance of 'Fitness that Fits' platform

favorite types of workout. In the following, our interest focuses on presenting the recommender model implemented in the platform.

4 RECOMMENDER MODEL

The model for the recommendation process underlying our platform is illustrated in Figure 2. It consists on a hybrid approach incorporating basic principles from content-based and neighbor-based collaborative filtering, on top of which three features are introduced:

(A) Identifying user preferences.
(B) Measuring diversity of recommendations.
(C) Diversity-aware replacement process.

These three novel features are labeled using (A), (B), (C) in Figure 2. The content-based and collaborative filtering steps are implemented upon basic approaches in the current prototype version of this work, and they can be seamlessly replaced by other existing methods with similar aim. We therefore concentrate the subsequent discussion on the three distinctive features listed previously.

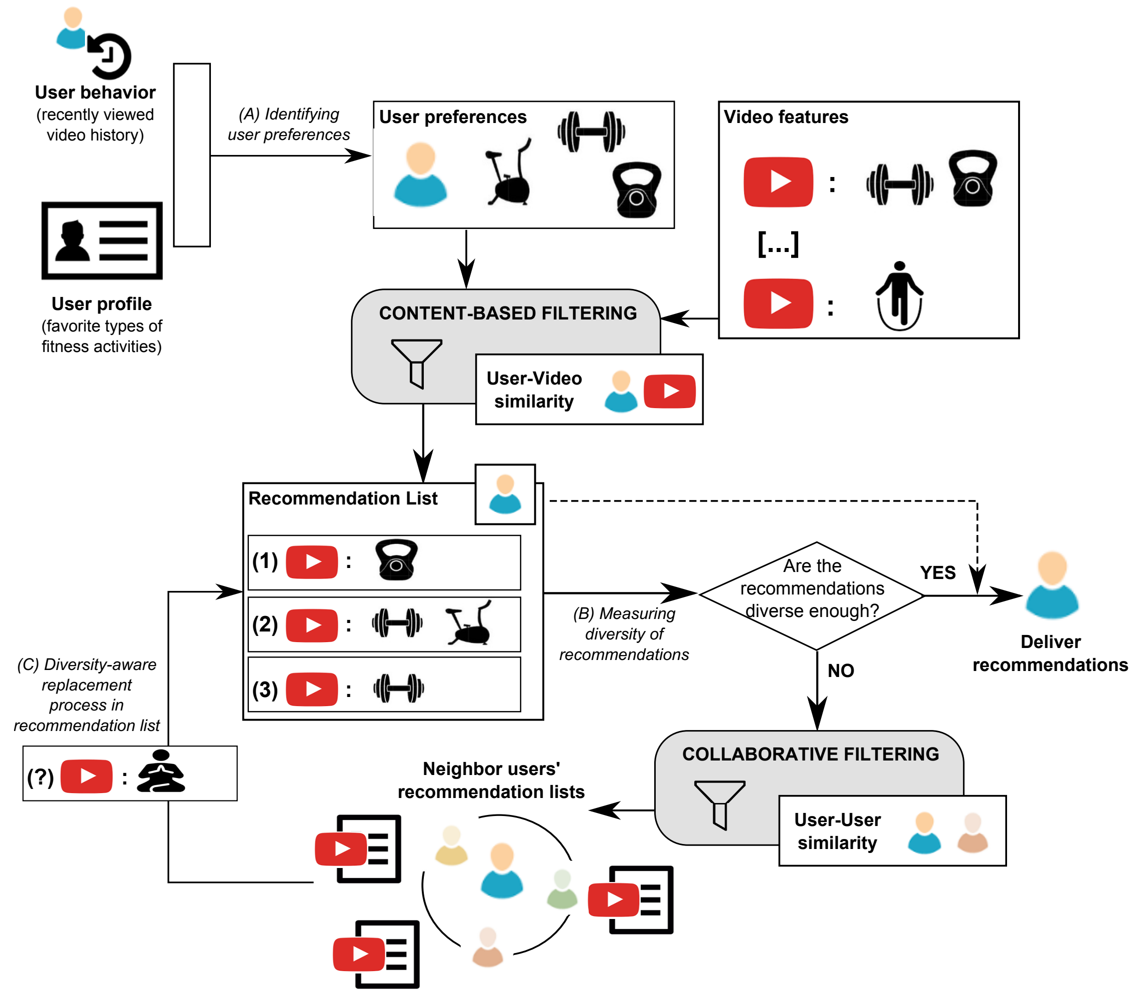

Figure 2: Scheme of the proposed model for workout video recommendation

4.1 (A) Identifying user preferences

Two sources of user data are taken as an input to model their current preferences: the user profile and the recent user behavior.

a) *User Profile*: A signed up user in the platform must specify their interests in different types of fitness activities (topics), by selecting her/his favorite ones. Let $m \gg 2$ be the number of topics available in the system[3]. Let $u_i \in U$ be the ith user. Her profile is modeled as an m-dimensional binary vector $P_i = (p_i^1, p_i^2, \ldots p_i^m)$, where:

$$p_i^j = \begin{cases} 1 & \text{if the } j\text{th topic is selected in } u_i \text{ profile,} \\ 0 & \text{otherwise.} \end{cases}$$

For instance, u_i profile could include "Kettlebells", "Aerobics" and "Pilates" as activated topics or favorite workout types.

b) *User Behavior*: This data is captured by analyzing the user's recent video viewing history of the user, namely the frequency at which the history contains videos labeled with each topic, e.g. "how often has u_i watched HIIT[4] videos over the last month?"

Both sources of information are combined into the current preferences of u_i. Let $\Gamma_i^j \geq 0$ denote the number of videos labeled with the jth topic in u_i recent history. Based on a normalized arctangent, a smoothed or "fuzzy" degree or preference $\hat{p}_i^j \in [0, 1]$ is calculated:

$$\hat{p}_i^j = arctan\left(\omega_i \cdot \frac{\Gamma_i^j}{\max_{k \in M} \Gamma_i^k} + (1 - \omega_i) \cdot p_i^j\right) \cdot \frac{4}{\pi} \qquad (1)$$

with $M = \{1, 2, \ldots, m\}$ the label index set. As shown in Eq. (1), $\hat{p}_i^j$ relies on a weighted average of (i) the relative frequency of workout videos containing topic j in the recent viewing history, and (ii) the (binary) profile information given by p_i^j. The weighting parameter $\omega_i \in [0, 1]$ describes the relative importance assigned to the user behavior information against her profile. It does neither require being fixed a priori nor it is equal for all users. Instead, ω_i

[3] For the current experimental version of the model, we have $m = 16$.

[4] High-Intensity Interval Training

is dynamically assigned to each user as follows:

$$\omega_i = min\left\{\frac{hl(i)}{2 \cdot \frac{1}{n}\sum_k hl(k)}\ ,\ 1\right\} \tag{2}$$

with $hl(i) \in \mathbb{N}$ the history length or number of recently viewed videos by u_i. This dynamic adjustment of ω_i *adapts* to the degree to which the user utilizes the system: the more engaged u_i has recently been (larger recent viewing history), the higher ω_i and the more relevance is assigned to the behavior information. If u_i recent engagement equals the average recent engagement of all users, i.e. if $hl(i) = \frac{1}{n}\sum_k hl(k)$, then $\omega_i = 0.5$.

As a result, a user's current preference vector $\hat{P}_i = (\hat{p}_i^1, \hat{p}_i^2, \ldots \hat{p}_i^m)$ is yielded for each $u_i \in U$. Based on the cosine similarity between a user preference and a video representation $P_v = (p_v^1, \ldots, p_v^m)$, $p_v^j \in \{0, 1\}$, a content-based filtering process is subsequently applied, leading to a preliminary video recommendation list of size N. This list might still be adjusted before being delivered to the end user, as explained in the following two subsections.

4.2 (B) Measuring diversity of recommendations

One of the aims of the proposed system is to provide users with recommended videos that are both relevant (in accordance with their current preferences) and diverse. Diversity in workout recommendations may not only help exploring "new" types of workout the user might potentially like, but also fosters variety of workouts in such recommendations to prevent an eventual sense of boredom. Let $R_i = (r_{vj}^i)_{N\times M}$ be a matrix representation of u_i's recommendation list, where the vth row contains the M-dimensional vector representation of the vth recommended video, hence $r_{vj}^i = 1$ if the vth recommended video is labeled with topic j, and $r_{vj}^i = 0$ otherwise. By introducing a diversity threshold $\delta \in]0, 1]$, the diversity level of R_i, denoted $\mathcal{D}(R_i)$, is measured and compared against δ, predicated on the number of topics appearing in at least one video in R_i ($\bigvee$ denotes the logical disjunction 'OR' operation):

$$\mathcal{D}(R_i) = \frac{\sum_j \left(\bigvee_{v=1}^{N} r_{vj}^i\right)}{M} \tag{3}$$

If $\mathcal{D}(R_i) \geq \delta$ (i.e. the ratio $\mathcal{D}(R_i)/\delta \geq 1$ as shown later on in experiments) then the recommended videos for u_i are sufficiently diverse and they are supplied to the user. Conversely, if $\mathcal{D}(R_i) < \delta$ then a neighborhood-based collaborative filtering approach is adopted to further diversify R_i based on the information extracted from similar users' recommendation lists. Importantly, we adopt a variant of classical user-user collaborative filtering in which, once the K most similar users to the target user have been identified (predicated on the similarity between their current preference vectors $\hat{P}_i$), we analyze those neighbor users' recommendation lists, as opposed to existing approaches that apply a rating prediction function.

4.3 (C) Diversity-aware replacement process

An iterative procedure is introduced to diversify the recommendation list R_i. The procedure is characterized by replacing - at each iteration - one of the videos recommended to the target user u_i with another video stemming from one of her neighbor users' recommendation list, based on the following steps:

(1) Sample one of the K neighbor users of u_i, denoted $u_{i'}$, by normalizing similarities $sim(u, u')$ into probabilities for each neighbor to be picked. Retrieve from the database the matrix $R_{i'}$ containing his latest list of recommendations.
(2) Check the rows (recommended videos) in $R_{i'}$ in descending order. Choose the first occurrence of video containing *at least* one topic j that does not appear in R_i. If no videos in $R_{i'}$ hold this condition, return to step (1).
(3) Assume that the selected video in $R_{i'}$ corresponds to the vth row of the neighbor's recommendation matrix. Then, in the target user matrix R_i, replace the existing video in the vth position with the selected video from $u_{i'}$ recommendations.

As a result of an iteration, a single-video replacement is made on R_i, after which its diversity level is measured again. The overall iterative process described in Section 4.2 and 4.3 is repeated until the recommendations are diverse enough, or an *a priori* fixed maximum number of iterations is exceeded. In either case, the final adjusted recommendation list is provided to u_i.

4.4 Preliminary Experiments

This subsection briefly outlines a preliminary evaluation conducted on the current version of 'Fitness that Fits' underlying model. We remark that despite the considerable volume of real labeled video data available, the present study relies exclusively on a small and partly synthetic dataset of user information (profiles and viewing histories). Deploying the platform into a Web environment and gathering larger amounts of user data for a comprehensive evaluation, constitutes our most immediate direction of future work.

We consider a sample of 10 users, a recommendation list size of $N = 30$, a number of neighbor users $K = 3$ for the diversification strategy, and a diversity threshold $\delta = 0.375$ (at least 6 out of 16 topics to appear in R_i). This preliminary experiment focuses on measuring the diversity ratio $\mathcal{D}(R_i)/\delta$ and the average similarity or *relevance* $\mathcal{S}(R_i)$ in the user's recommendation matrix, *before* and *after* applying some replacements under the proposed diversification strategy: until δ is achieved or at most $N/2$ replacements are made on R_i. The similarity score $\mathcal{S}(R_i) \in [0, 1]$ is calculated at the average of cosine similarities between the user current preferences and her N recommended video representations. Figure 3 summarizes the results obtained for the sample considered. Intuitively, for those users whose preliminary recommendations were already sufficiently diverse (in the Example, u_7 and u_{10}) no changes are required on the preliminary recommendations. The initial and final values of $\mathcal{D}(R_i)$ (resp. $\mathcal{S}(R_i)$) are shown above (resp. below) the plot bars.

This initial evaluation shows that, although the model can effectively increase the diversity of recommended workout videos, this normally comes at the expense of a decline in the relevance (similarity) or such recommendations with respect to the target user preferences. This is not surprising, since the initial content-based filtering stage (before diversifying) relies exclusively on the users' preferences, and the current use of a cosine similarity may make the resulting relevance sensitive to any zeros in either user preferences of video representations. Moreover, data containing recent viewing

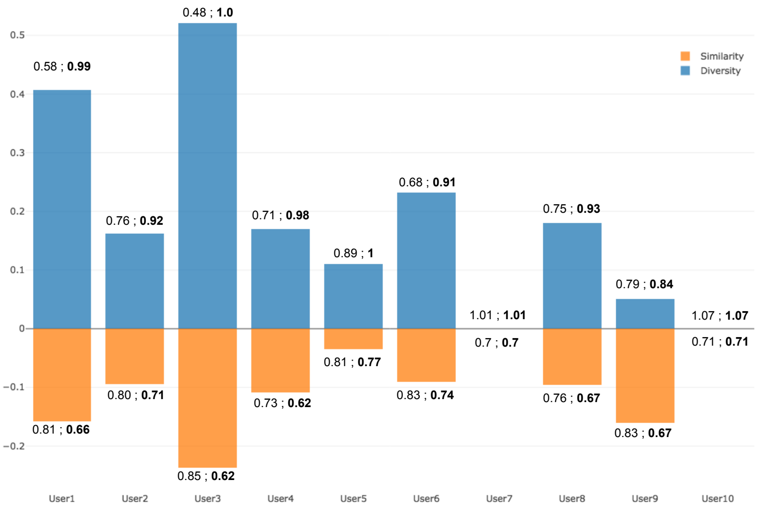

Figure 3: Variation in the diversity ratio ($\mathcal{D}(R_i)/\delta$) and relevance (average similarity to the user, $\mathcal{S}(R_i)$) of R_i for $\delta = 0.375$

histories is rather scarce in the present prototype, therefore the weighting parameter ω_i tends to be low for most users and their static profile information (favorite topics) is prioritized. We argue that an online evaluation of users' experience with the system, for instance by tracking clicks on recommended videos and analyzing whether such clicks have been predominantly on "replacement" videos or not, will be an interesting direction to extend the proposed approach into a more adaptive one, where depending on the user's response towards diversity, a tailored trade-off between diversity and relevance is sought for her/him. Another interesting finding is the dramatic increase in the diversity of R_1 and R_3 with respect to other users. This is largely due to some videos in the system having associated more topic labels than others.

5 CONCLUSION AND FUTURE DIRECTIONS

This contribution presented 'Fitness that Fits', a prototype platform for recommending physical workout videos upon labeled video data from the Youtube-8M dataset. Besides integrating basic content-based and collaborative filtering mechanisms, the proposed recommender model incorporates novel features for the flexible modeling of user preferences based on their profile and recent viewing behavior. Furthermore, an iterative replacement strategy inspired by neighborhood-collaborative filtering is introduced to promote diversified recommendation lists for users to enhance with different types of fitness activities.

Besides the platform deployment, subsequent acquisition of more real user data and the elaboration of a complete experimental study, we also postulate the following directions for future research:

- Recent efforts have been put in personalization services for promoting healthy habits, e.g. via food recommendations for positive nourishment practices [13], hence we aim at incorporating labeled videos on healthy eating habits, and investigating the joint use of fitness and healthy eating user/content data into video recommendations for healthier living.
- Recommend longer (composite) workout recommendations by producing sequences of smaller workout videos while advocating diversity in such workouts.
- Incorporate additional types of explicit and implicit preferences from real users, e.g. liked-disliked videos from Youtube and favorite videos.

ACKNOWLEDGMENTS

The authors would like to thank anonymous reviewers for their insightful and constructive comments on the initial stage of this research, some of which would lay the foundations for future work.

REFERENCES

[1] J. Berndsen, A. Lawlor, and B. Smyth. 2017. Running with Recommendation. In *Proc. 2nd International Workshop on Health Recommender Systems; 11th International Conference on Recommender Systems (RecSys 2017)*. 18–21.

[2] S. Borreani. Last access: 24th June 2018. The Fitness Sector in the Internet: Missed Opportunities. *Source: https://gymfactory.net* (Last access: 24th June 2018).

[3] G. M. Cerón-Rios, D. M. Lopez Gutierrez, B. Díaz-Agudo, and J. A. Recio-García. 2017. Recommendation System based on CBR algorithm for the Promotion of Healthier Habits. In *Proceedings of ICCBR 2017 Workshops (CAW, CBRDL, PO-CBR), Doctoral Consortium, and Competitions co-located with the 25th International Conference on Case-Based Reasoning (ICCBR 2017), Trondheim, Norway, June 26-28, 2017*. 167–176.

[4] E. Chen. 2017. Youtube-8M Video Understanding Challenge Approach and Applications. In *CVPR'17 Workshop on YouTube-8M Large-Scale Video Understanding*.

[5] P. Covington, J. Adams, and E. Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. ACM, New York, NY, USA, 191–198.

[6] J. Davidson, B. Liebald, J. Liu, P. Nandy, Taylor Van V., U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston, and D. Sampath. 2010. The YouTube Video Recommendation System. In *Proceedings of the Fourth ACM Conference on Recommender Systems (RecSys '10)*. 293–296.

[7] S. Dharia, M. Eirinaki, V. Jain, J. Patel, I. Varlamis, J. Vora, and R. Yamauchi. 2018. Social recommendations for personalized fitness assistance. *Personal and Ubiquitous Computing* 22, 2 (01 Apr 2018), 245–257.

[8] D. Elsweiler, B. Ludwig, A. Said, H. Schaefer, and C. Trattner. 2016. Engendering Health with Recommender Systems. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. 409–410.

[9] R. Harakawa, T. Ogawa, and M. Haseyama. 2017. Extracting Hierarchical Structure of Web Video Groups Based on Sentiment-Aware Signed Network Analysis. *IEEE Access* 5 (2017), 16963–16973.

[10] E. T. Luhanga, A. A. E. Hippocrate, H. Suwa, Y. Arakawa, and K. Yasumoto. 2018. Identifying and Evaluating User Requirements for Smartphone Group Fitness Applications. *IEEE Access* 6 (2018), 3256–3269.

[11] C.L. Sánchez-Bocanegra, J.L. Sevillano-Ramos, C. Rizo, A. Civit, and L. Fernandez-Luque. 2017. HealthRecSys: A semantic content-based recommender system to complement health videos. *BMC Medical Informatics and Decision Making* 17, 63 (2017), 1–10.

[12] F. Scheidegger, L. Cavigelli, M. Schaffner, A. C. I. Malossi, C. Bekas, and L. Benini. 2017. Impact of temporal subsampling on accuracy and performance in practical video classification. In *2017 25th European Signal Processing Conference (EUSIPCO)*. 996–1000.

[13] C. Trattner and D. Elsweiler. 2017. Food Recommender Systems: Important Contributions, Challenges and Future Research Directions. *Collaborative Recommendations: Algorithms, Practical Challenges and Applications (World Scientific)* abs/1711.02760 (2017).

[14] P. Xie, R. Salakhutdinov, L. Mou, and E. P. Xing. 2017. Deep Determinantal Point Process for Large-Scale Multi-label Classification. In *2017 IEEE International Conference on Computer Vision (ICCV)*. 473–482.