DATA-DRIVEN SONG RECOGNITION ESTIMATION USING COLLECTIVE MEMORY DYNAMICS MODELS

Christos Koutlis[1] **Manos Schinas**[1] **Vasiliki Gkatziaki**[1]
Symeon Papadopoulos[1] **Yiannis Kompatsiaris**[1]
[1] CERTH-ITI, Thessaloniki, Greece
ckoutlis@iti.gr, manosetro@iti.gr, vasgat@iti.gr, papadop@iti.gr, ikom@iti.gr

ABSTRACT

Cultural products such as music tracks intend to be appreciated and recognized by a portion of the audience. However, no matter how highly recognized a song might be at the beginning of its life, its recognition will inevitably and progressively decay. The mechanism that governs this decreasing trajectory could be modelled as a forgetting curve or a collective memory decay process. Here, we propose a composite model, termed T-REC, that involves chart data, YouTube views, Spotify popularity of tracks and forgetting curve dynamics with the purpose of estimating song recognition levels. We also present a comparative study, involving state-of-the-art and baseline models based on ground truth data from a survey that we conducted regarding the recognition level of 100 songs in Sweden. Our method is found to perform best among this ensemble of models. A remarkable finding of our study pertains to the role of the number of weeks a song remains in the charts, which is found to be a major factor for the accurate estimation of the song recognition level.

1. INTRODUCTION

Music is a form of art that attracts the vast majority of global population and has a remarkable impact on people's emotions and behavior. In particular, in-store consumer purchase behavior has been related to background music in the research literature [9, 21, 22, 31]. Also, the role of music popularity, liking and recognition levels in shopping intentions [4, 35] and the perception of time [2] has been investigated. Background music providers supply companies with music playlists with the purpose of optimizing the in-store experience of their customers and their brand perception. Having effective means of estimating song recognition can provide such companies with a useful tool for generating better playlists. Motivated by the above, in this paper we propose an accurate song recognition model as basis for experimentally measuring the impact of song recognition on in-store purchase behavior. [1]

Although song *recognition* is the focus of this paper, song *popularity* is more frequently encountered in the research literature. The popularity of songs is a concept used to express how much attention a certain song *currently* receives. There have been attempts towards determining song popularity making use of the online available information from posts in microblog websites [12, 23, 28, 29] and in the blogosphere [1], search queries and number of shared files in peer-to-peer networks [17, 28], play counts in social media music sites such as Last.fm [3, 29], the amount of time of radio play, the music industry awards that it received [25] and popularity indices provided by streaming platforms such as Spotify [3]. Of course the traditional ways of determining music popularity such as the Billboard Magazine chart are also used for comparison with the modern web-based popularity indices [16, 17].

Music exhibits its own complex dynamics in terms of popularity growth and decay while the means used to promote it in the public constantly evolve and capitalize on the advances and trends of digital media and communication technologies. During the last decade, researchers have investigated special attributes of songs that may lead to a successful release [8, 15, 36], and have attempted to successfully predict hit songs [6, 16]. However, while song popularity is a research topic that has attracted intense academic interest the level of a music track's recognition is a notion that has been significantly less studied.

As song recognition we define the fraction of the audience that recognizes (comprehend that they have heard it before) a specific music track through audio exposure. This notion is different from the notion of song popularity as a song might no longer be trending (for instance an old song no longer placed in the charts) but at the same time a considerable portion of the music audience might recognize its tune. To further illustrate this differentiation, in Table 1 we present the most popular songs of 2018 [2] and most recognized songs of all time [3]. It is apparent that songs on the left column currently overwhelm the charts and online playlists, but songs on the right column are

[1] We consider atmospherics, such as high/low recognition music, as causal factors for consumer behavior in stores.

[2] According to Spotify's "Top Tracks of 2018" list.

[3] According to experimental results obtained by "Hooked On Music" (http://www.hookedonmusic.org.uk/) and published on BBC's webpage (https://www.bbc.com/news/science-environment-29847739).

Top 2018 songs	Top recognized songs
God's plan/Drake	Wannabe/Spice Girls
SAD!/XXXTENTACION	Mambo No 5/Lou Bega
rockstar/Post Malone	Eye Of The Tiger/Survivor
Psycho/Post Malone	Just Dance/Lady Gaga
In My Feelings/Drake	SOS/ABBA
Better Now/Post Malone	Pretty Woman/Roy Orbison
I like It/Cardi B	Beat It/Michael Jackson
One Kiss/Calvin Harris	I Will Always Love You/Whitney Houston
IDGAF/Dua Lipa	Don't You Want Me/The Human League
FRIENDS/Marshmello	I Don't Want To Miss A Thing/Aerosmith

Table 1: Left column: top 2018 songs in terms of popularity. Right column: top recognized songs of all time.

surely recognized by a very high percentage of the population even though they are not currently popular.

To estimate the portion of the audience that recognize a music track, one should take into account the cognitive aspects of the problem. That is to say, collective memory dynamics and more precisely the mechanisms that govern its initial increase and its decay after the initial period of popularity. For the more general concept of human memory decay many studies have been conducted [10,11,20,24] indicating forgetting curves with exponential decay dynamics. Music-specific research has adopted exponential forgetting curves [7, 14] for song "freshness" assessment, as well. A method with double exponential dynamics was proposed as a general memory decay model [5], while log-normal dynamics were employed to model the dynamics of scientific paper impact [33]. Equally important is to take into account the notions of learning curves [26] and over-learning [27] in order to determine the initial amount of learned information and the velocity of forgetting. In many studies researchers argue for the significant role of (i) repetition of a stimulus in learning [10, 13, 19, 26, 27] and (ii) the degree of original learning in the velocity of forgetting [10, 20, 30, 34]. Namely, the more exposed to a stimulus humans are the higher the initial amount of learned information is and the slower they forget it.

Here, we propose T-REC, a song recognition model that takes as input the chart positions a track has gained along with the respective dates, its current YouTube views and Spotify popularity. T-REC also considers sigmoid learning curve dynamics, exponential decay forgetting curve dynamics and a decay rate being a function of the number of weeks each track is maintained in the charts, which we consider here as a proxy of the original learning degree. In other words, the number-of-weeks feature is an indicator of how strongly the audience is exposed to a specific tune and as it increases, the forgetting procedure (i) starts from a higher point and (ii) decelerates further, as indicated by the related research literature on human memory. Eventually, T-REC results in the estimation of song recognition levels per market and globally. Other competitive models are also considered for comparison purposes. We have conducted a survey for the estimation of the current level of recognition of 100 songs in Sweden, which we then use as ground truth for the evaluation of our method's performance and for comparison with the other methods. We make the resulting data available for the community [18].

2. MATERIALS AND METHODS

2.1 Data

For estimating the recognition levels of music tracks, our starting point was a list of tracks provided by *Soundtrack Your Brand*, a collaborating background music provider. The list consists of 39,466 tracks from 21,450 artists and from 75 countries. We also make use of data from 211 charts, 198 track charts and 13 singles charts, that span long periods of time (in some cases from the 60s until today) from 62 countries around the globe including Sweden. We also used the Spotify API to annotate chart entries with the Spotify id and International Standard Recording Codes (ISRC) [4] of each of the songs. Since our user study was carried out on a Swedish population, we present the monitored charts for Sweden along with the corresponding monitored periods in Table 2.

chart name	since	until
Spotify Daily Chart	2017-01-01*	2018-03-06
Spotify Weekly Chart	2016-12-23	2018-03-01
Veckolista Svenskt Topp-20	2015-01-17	2018-06-15
Veckolista Singlar	1988-01-16	2018-06-15
Veckolista Heatseeker	2015-01-10	2018-06-15
Veckolista Svenska Singlar	2015-01-10	2015-01-16
SINGLES TOP 100	1975-11-08	2018-06-01
Sweden Top 20	2001-06-12	2018-07-07
Sweden Singles Top 100	2017-12-29	2018-07-05

Table 2: The list of Swedish charts we used in this study. The first column presents the chart name, the second and third columns present the start and end dates of monitoring respectively. *All dates are in YY-MM-DD format.

Most songs do not make it in the charts, thus we additionally employ YouTube views and Spotify popularity of tracks as current track popularity proxies. Knowing the Spotify id of the tracks and the id of an associated official video in YouTube [5], we retrieved these two signals by using the public APIs offered by Spotify and YouTube respectively. The intuition behind the use of these two metrics, is that they reflect the exposure of a song in two widely used platforms. Number of video views in YouTube is a direct measure of how many people heard a song. On the other hand, although Spotify popularity is a score generated internally by Spotify and the exact formula is not known, that score reflects the actual number of streams a song received recently. Therefore, we can safely assume that a song having a high popularity score is currently listened more than songs with a lower score.

2.2 Models

2.2.1 T-REC

The proposed song recognition model builds upon three main components, the *recognition growth* that represents the level of recognition a track reaches during its initial

[4] https://isrc.ifpi.org/

[5] To this end, we used the Soundiiz (www.soundiiz.com) service, which supports playlist conversion between platforms.

prosperity time (when it is placed in charts), the *recognition decay* that represents the collective memory decay process (i.e. the mechanism of collective forgetting of songs) and the *recognition proxy-based adjustment* that adjusts the recognition level of tracks, which is especially useful for tracks with no chart information.

Having annotated chart entries with the corresponding ISRC, we were able to retrieve the positions of tracks in the Swedish charts of Table 2. These are then used to estimate their *recognition growth* (in Sweden) according to Equation 1:

$$g(t) = \begin{cases} 100 \cdot \frac{c_K+1-r_K(t)}{c_K} \cdot \sigma_1, \\ \qquad\qquad\qquad \text{if track in chart } K \text{ at time } t \\ 0, \text{ otherwise} \end{cases} \tag{1}$$

where c_K is the number of tracks in chart K, $r_K(t)$ is the rank of the track in chart K at time $t \in [t_0, t_{today}]$ [6] and $\sigma_1 = \sigma_1(\theta_0, \theta_1, x) = (1 + e^{-\theta_1 \cdot x + \theta_0})^{-1}$ adjusts the rank's importance using an S-shaped learning curve with $x \in (0, +\infty)$ and $\theta_0, \theta_1 \in \mathbb{R}$. The logistic part of the model is incorporated to control the importance of a chart position given the number of weeks x the track has remained in the charts. If a track remained in the charts for only one week its rank's importance would be a lot lower (54.9%) compared to it remaining for 20 weeks (98.2%). Therefore, in the first case the decay process will begin from a much lower point. The value of $g(t)$ is assigned to all $g(t_i)$ with $t_i \in [t-n+1, t]$ according to the chart's frequency e.g. if it is weekly $n = 7$. If a track gains multiple values at a single date, the maximum value is used.

Towards formally defining the *recognition decay*, we build on findings from research literature in the area of human memory, and more precisely on the concept of forgetting curves. A forgetting curve is the rule by which the memory regarding a specific learned item is reduced. In our case we consider as learned items the music tracks. Hence, we aim at estimating the function that describes the forgetting procedure that has been proposed to be exponentially decreasing in many studies [20, 24]. We also opt for the exponentially decreasing forgetting curve for song recognition but at a less steep rate. It is natural to consider that the level of *recognition decay* is impossible to be higher than the level of *recognition growth* at its peak for a particular track. Also, each time the track reemerges in the charts the forgetting procedure restarts from a new higher point of recognition. Additionally, we consider a variable decay rate as a reasonable consideration would be that not all songs' recognition decays with the same velocity. The *recognition decay* is defined in Equation 2:

$$d(t) = \begin{cases} g(t), \text{ if } d(t-1) \leq g(t) \\ \sigma_2 \cdot d(t-1) + (1-\sigma_2) \cdot g(t), \text{ otherwise} \end{cases} \tag{2}$$

where $\sigma_2 = \sigma_2(\phi_0, \phi_1, x) = (1 + e^{-\phi_1 \cdot x + \phi_0})^{-1}$ is the recognition retention percentage with x being the number of weeks the track has remained in the charts and $g(t)$ is the previously defined in Equation 1 *recognition growth*. If a track has remained for a long time in the charts its retention percentage would be considerably high and its forgetting process would be rather slow, while if a track has remained in the charts for only few weeks its retention percentage would be low and its forgetting process fast.

The first logistic function σ_1 controls the initial recognition level from which the decreasing trajectory begins and the second logistic function σ_2 controls the velocity of recognition decay, both individually per track.

To model the *recognition proxy-based adjustment*, we consider a multiple linear regression model with input the track's current Spotify popularity index (P_S) and the log-transformed YouTube views (P_{YT}) as in Equation 3:

$$s(t_{today}) = \alpha_0 + \alpha_1 \cdot log(P_{YT}) + \alpha_2 \cdot P_S \tag{3}$$

The composite T-REC model is defined as a linear combination of *recognition decay* and *recognition proxy-based adjustment* at t_{today}:

$$\text{T-REC} = w_0 + w_1 \cdot d(t_{today}) + w_2 \cdot s(t_{today}) \tag{4}$$

2.2.2 Competitive Models

In order to perform a comparative study, four competitive models are employed for the task of song recognition estimation. Two of them are well-known regression models, one is related to collective memory decay while the last one is the plain Spotify popularity index (P_S).

The first model is based on Multiple Linear Regression (MLR) and the second on Random Forests (RF). The log-transformed YouTube views and the Spotify track popularity are considered as inputs to these models and actual recognition as their target. MLR actually corresponds to the proxy-based adjustment introduced in Equation 3. The third competitive model is the state-of-the-art log-normal decay model (LOGN) [33] that Wang et al. developed for modelling the decay process of scientific paper citations. We use the form of this model that is presented in the supplementary material of [5], namely Equation 5:

$$l(t) = e^{\left[ln\left(\frac{\lambda}{\sqrt{2\pi}\sigma}(c^t+m)\right)-\mu^2\right]} \cdot t^{\frac{\mu}{\sigma^2}-1} \cdot e^{-\frac{ln^2(t)}{2\sigma^2}} \tag{5}$$

where t is time, $c^t = m\left[e^{\lambda\Phi\left(\frac{ln(t)-\mu}{\sigma}\right)} - 1\right]$, $\Phi(\cdot)$ is the cumulative distribution function of the normal distribution and λ, μ, σ, m are arbitrary parameters.

2.2.3 Optimization and Evaluation

We consider a holdout strategy (70% training, 30% test) for the models' evaluation as described in [32]. The optimization of all models' parameters is performed in the training set by the truncated Newton algorithm as implemented by the SciPy package. We use as objective function the mean absolute error between measured (by the user study) and computed (by each model) song recognition.

[6] As t_0 we set the song's release date and as t_{today} the current date.

	class	#	fraction	Sweden
gender	male	521	50.05%	50.24%
	female	520	49.95%	49.76%
age	18-24	54	5.19%	18.15%*
	25-34	422	40.54%	22.46%
	35-44	277	26.61%	20.01%
	45-54	244	23.44%	21.12%
	55-65	44	4.22%	18.26%

Table 3: Demographics of the test population and Sweden (normalized within the group of people between 15 and 65 years old). *This figure refers to 15-24 age group.

The models' performance is then evaluated in the test set by the mean absolute error (MAE) between the measured and the computed recognition as in Equation 6:

$$\text{MAE} = \frac{\sum_{i=1}^{k} \mid x_i - y_i \mid}{k} \tag{6}$$

where k is the number of tracks, x_i is the measured recognition for track i and y_i is the computed recognition for track i. A perfectly accurate model would lead to a MAE value of 0.

2.3 User study

To proceed with the user study, we employed an initial and much simpler version of the recognition score. This initial version had a constant decay rate across all tracks and for the tracks with no chart data the average recognition score of the closest, in terms of YouTube views and Spotify popularity, tracks was considered as their recognition score. [7]

After the assignment of the initial recognition score (corresponding to the time the survey was conducted) to each of the 39,466 tracks, we formed two lists. One list containing the 600 most recognized tracks in Sweden and a second containing the 600 least recognized tracks in Sweden. [8] Consequently, 50 tracks were randomly chosen out of each of these two lists as representative of high and low recognition tracks.

A study was then conducted in order to obtain the actual recognition percentages for each of these 100 songs among a test population of 1041 annotators in Sweden. [9] We divided the initial list of 100 songs in 10 groups of 10 songs (5 of low and 5 of high recognition level in a randomized order), then each participant listened to 30-second samples of all the songs of one group and for each song he/she indicated whether he/she recognized it or not. We had $\sim$100 respondents per song [10] so we got a score 0-100 based on the percentage of respondents who responded positively.

[7] The rationale behind the alterations on this model that led to T-REC is illustrated in the results section.

[8] Given that recognition estimation is the result of a sampling process, we expect measurements in the extremes (i.e. least and most recognised songs) to be less noisy than in intermediate recognition levels. This motivated our choice to perform the initial song selection out of two distinct sets (high, low).

[9] The study was performed through the Cint survey platform (`https://www.cint.com/`).

[10] Some variability was due to the fact that not all respondents completed the process successfully.

θ_0	θ_1	ϕ_0	ϕ_1	α_0
0.233	0.043	0.847	0.029	1.299
α_1	α_2	w_0	w_1	w_2
0.999	-0.093	22.586	0.452	0.928

Table 4: Parameter values for T-REC after fitting.

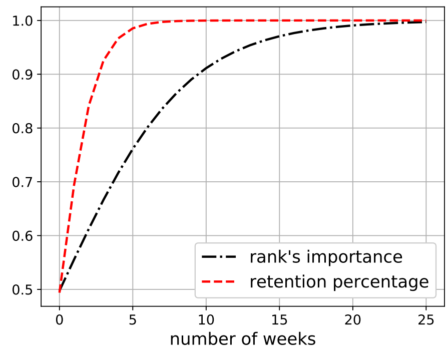


Figure 1: The logistic parts of *recognition growth* (rank's importance) and *recognition decay* (retention percentage) components as formed after the model fitting.

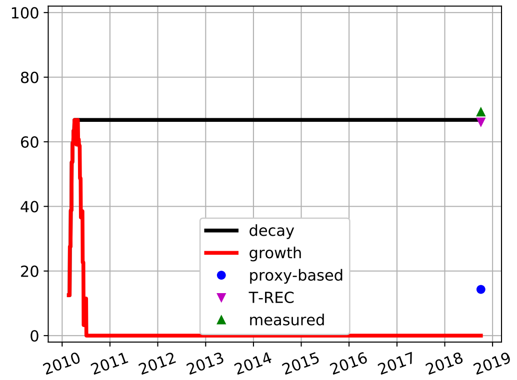
We consider the recorded responses as ground truth for our experiments and we evaluate our model as well as the competitive models on this basis. Demographics of the test and Swedish population [11] are illustrated in Table 3. We observe divergent age demographics, yet almost identical gender demographics between the test and actual population. As the selection of annotators was carried out by Cint, we could not better approximate the Swedish population distribution. Despite the over-representation of some age groups and under-representation of others, T-REC is still a sound methodology; given a different population sample to learn from, the model tuning step (section 2.2.3) would lead to a slightly different recognition estimation model.

3. RESULTS

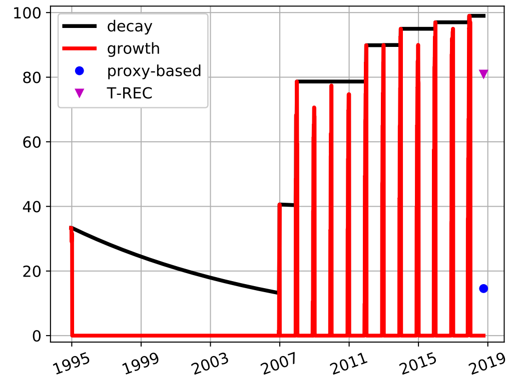
The analysis of the survey data shows that the initial recognition score classified the tracks effectively with 50/50 (100%) correctly labeled as low and 37/50 (74%) correctly labeled as high recognition (measured recognition <50% is considered as low, while >50% as high). The 13 songs that were falsely classified as high recognition actually obtain a smaller recognition score than the rest (on average 6 units lower). Despite the promising classification performance the measured recognition was in many cases far from the computed score especially in cases of tracks with no chart data. Thus, we developed the updated version of the recognition score (T-REC) described in section 2.2.1.

Table 4 presents the parameter values of T-REC after optimization. The model gives significant weights on both *recognition decay* (w_1) and *recognition proxy-based adjustment* (w_2) components, but considers YouTube views

[11] Sources: statista.com/statistics/521717/sweden-population-by-age/, statista.com/statistics/521540/sweden-population-by-gender/



(a) Rude Boy - Rihanna



(b) All I Want for Christmas Is You - Mariah Carey

Figure 2: T-REC components (*recognition growth*, *decay* and *proxy-based adjustment*) for two highly recognized songs.

(α_1) as more important than Spotify popularity (α_2) for the under study problem. The impact of the rest of the parameters on the final model, namely the shape of the two logistic functions that control the *recognition growth* (θ_0, θ_1) and *recognition decay* (ϕ_0, ϕ_1) components is illustrated in Figure 1. The logistic part of *recognition growth* (rank's importance) is less steep than the logistic part of *recognition decay* (retention percentage), indicating that a music track will need almost 7 weeks in the charts to achieve a very slow rate towards oblivion, but at least 25 weeks to achieve its highest contemporary recognition.

Moreover, two examples of how T-REC models the mechanism of song recognition decay are illustrated in Figure 2. As illustrated in Figure 2a, the song "Rude Boy" by Rihanna stayed in Swedish charts for 19 weeks, and according to the *recognition decay* component it maintained 99.9% of its initial recognition. Consequently, the *recognition proxy-based adjustment* input adjusts T-REC very close to the measured recognition (error=3.11). A different example presented in Figure 2b shows that Mariah Carey's "All I Want for Christmas Is You" initially was not a big hit in Sweden, remaining for only three weeks in the charts back in 1995. Afterwards, its recognition exhibited a significant decrease during the next decade, but after 2007 when the song kept reemerging in the charts every year its recognition decay rate slowed down and both its *recognition growth* and *decay* components grew larger. The *recognition proxy-based adjustment* component adjusts T-REC a little lower. Although we lack ground truth for this song to compare it to T-REC's estimation (as it was not in the survey's lists), we believe that 80.99% recognition is closer to the real recognition rate [12] than the 98.99% computed by the *recognition decay* component, which is obviously too high even for a massive hit such as this.

Figure 3 illustrates the performance of T-REC on estimating the actual current recognition level of songs in Sweden. Most of the points are concentrated close to the identity line except for some tracks of intermediate recognition levels, which are overestimated. Table 5 compares the performance of T-REC with the one of all competitive models in terms of average MAE. All models (except for Spotify popularity index) are trained on 100 different training sets, each containing 70 randomly selected tracks out of the initial set of 100 tracks and then their MAE is measured on the 100 corresponding test sets, each containing the remaining 30 tracks. T-REC exhibits the best performance among all models with a very high statistical significance level as indicated by the p-value=10^{-17}, according to the Wilcoxon signed rank paired test. [13]

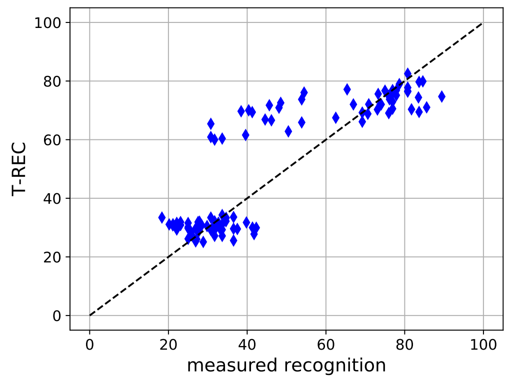


Figure 3: Scatter plot with the T-REC recognition score on the y axis and measured recognition on the x axis.

As additional information we provide long lists of Top-100 recognized songs and the corresponding Top-10 artists (according to T-REC) for Sweden and USA, in the supplementary material. Furthermore, we compare T-REC's Top-N lists with Billboard's "The Hot 100's All-Time Top 100 Songs" list.The results show that T-REC assigns top scores to most of these songs as well. This fact also holds (but to a lower degree) when the chart data from "The Hot 100" are omitted from the input list of charts.

The diverging behaviour of the songs with intermediate recognition level in Figure 3 is also apparent in YouTube

[12] Or only slightly underestimating it given that the top-3 measured recognition percentages of our survey are 89.42, 85.57 and 84.61.

[13] This is the maximum p-value among all four comparisons.

model	AMAE
P_S	20.63
MLR (Equation 3)	11.30
RF	10.27
LOGN (Wang et al. 2013 [33])	22.00
T-REC	**8.50**

Table 5: Average MAE over 100 randomly selected test sets for T-REC, Spotify popularity (P_S), Multiple Linear Regression (MLR), Random Forest (RF) and log-normal (LOGN) models. For Spotify popularity we computed once the mean absolute error over all 100 tracks.

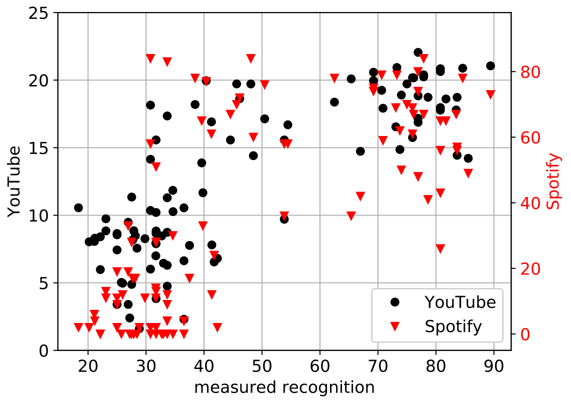


Figure 4: Scatter plot - y axes: YouTube views (log) and Spotify popularity, x axis: measured recognition.

and Spotify indices as shown in Figure 4. One possible explanation for this behavior is that 15 out of these 17 songs have been released during the last three years and they still are in their initial popularity phase. Thus, there has not passed a considerable amount of time in order for these tracks to experience significant recognition decay, which T-REC would likely capture. As exemplified in Figure 5 the more recent the track the bigger the error our model produces, which is a limitation of the proposed model, even though the average errors in all periods are small (the maximum is 9.5) and less than any other compared model.

Finally, we would like to elaborate on the rationale behind the refinements that we performed on our model in order to take its final form (Equation 4). In Figure 4 a linear interaction is observed between (i) the log-transformed YouTube views and Spotify popularity and (ii) the measured recognition, with Pearson correlation coefficients 0.79 and 0.71 respectively. Thus, we incorporated the multiple linear model with the corresponding input quantities as *recognition proxy-based adjustment* component in the final T-REC formula. The consideration of a constant decay rate in the formula of *recognition decay* is not plausible, since it would further lead to a zero rate as the best choice after model fitting, which is highly unrealistic, as the model would degenerate into the *recognition growth* component. As a result, the final form of T-REC includes a variable decay rate across music tracks that depends on the number of weeks the track has remained in the charts. This refinement resulted in significantly lower errors showcasing the major role of the number-of-weeks feature in song recognition estimation. More specifically, the initial recognition score achieved a MAE of 12.32, while T-REC a much lower MAE of 8.50 as presented in Table 5.

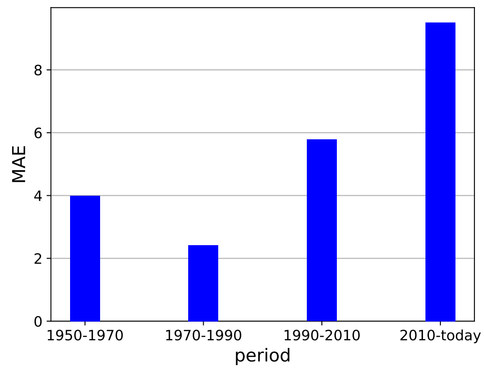


Figure 5: Mean absolute error of T-REC on tracks released in different periods of time.

4. CONCLUSIONS

In this work, we studied collective memory dynamics with regards to song recognition. We proposed a model for the approximation of the corresponding decreasing trajectory and the estimation of the current song recognition level. Our recognition model comprises three main components: a) growth, b) decay, and c) proxy-based adjustment and it leverages chart data, YouTube views, Spotify popularity and forgetting curve dynamics. Also, our method considers different recognition decay rates and initial recognition levels per song, according to the number of weeks the song has remained in the charts.

We compared our model to other state-of-the art and baseline models on the task of accurately estimating the current recognition level of songs. To this end, we conducted a study in Sweden in order to measure the recognition level of 100 songs, which we then used as ground truth for the models' evaluation. The experimental results showed that our method exhibits great performance on this task, much better than the competitive models with a high statistical significance level.

Finally, we reached two remarkable conclusions:

1. according to our model's parameters, a music song needs almost 7 weeks in the charts to achieve a very slow velocity towards oblivion and at least 25 weeks to achieve its highest contemporary recognition;
2. the role of the number-of-weeks feature incorporated in our model through the logistic functions is found to be of utmost importance for the accurate estimation of a song's recognition level.

Future work will include extensions that alleviate the deviation of recent tracks' recognition estimation and also account for demographic-specific estimations.

5. ACKNOWLEDGMENTS

This work is funded by the European Commission under the contract number H2020-761634 FuturePulse. We would also like to thank our collaborators from Soundtrack Your Brand, Magnus Rydén, Jasmine Moradi and Daniel Johansson, who contributed to the problem formulation, provided the initial list of tracks, and carried out the crowd sourcing experiment to measure the recognition rates that were used as "ground truth" in our experiments.

6. REFERENCES

[1] F. Abel, E. Diaz-Aviles, N. Henze, D. Krause, and P. Siehndel. Analyzing the blogosphere for predicting the success of music and movie products. In *2010 International Conference on Advances in Social Networks Analysis and Mining*, pages 276–280, 2010.

[2] N. Bailey and C. S. Areni. When a few minutes sound like a lifetime: Does atmospheric music expand or contract perceived time? *Journal of Retailing*, 82(3):189–202, 2006.

[3] A. Bellogín, A. P. de Vries, and J. He. Artist popularity: Do web and social music services agree? In *Proc. of the 7th International AAAI Conference on Weblogs and Social Media*, pages 673–676, 2013.

[4] G. Broekemier, R. Marquardt, and J. Gentry. An exploration of happy/sad and liked/disliked music effects on shopping intentions in a women's clothing store service setting. *Journal of Services Marketing*, 22:59–67, 2008.

[5] C. Candia, C. Jara-Figueroa, C. Rodriguez-Sickert, A. L. Barabási, and C. A. Hidalgo. The universal decay of collective memory and attention. *Nature Human Behaviour*, 3(1):82–91, 2019.

[6] C. Chiru and O. G. Popescu. Automatically determining the popularity of a song. In *International Joint Conference on Rough Sets (IJCRS 2017)*, pages 392–406, 2017.

[7] A. Chmiel and E. Schubert. Using psychological principles of memory storage and preference to improve music recommendation systems. *Leonardo Music Journal*, 28:77–81, 2018.

[8] P. François and R. Pierre. Hit song science is not yet a science. In *Proc. of ISMIR 2008*, pages 355–360, 2008.

[9] S. O. Daunfeldt, N. Rudholm, and H. Sporre. Effects of brand-fit music on consumer behavior: A field experiment. *HUI Working Papers 121, HUI Research*, 2017.

[10] H. Ebbinghaus. *Memory: A Contribution to Experimental Psychology*. Teachers College, Columbia University, New York, 1913.

[11] G. Elliott, C. L. Isaac, and N. Muhlert. Measuring forgetting: a critical review of accelerated long-term forgetting studies. *Cortex*, 54(100):16–32, 2014.

[12] J. Grace, D. Gruhl, K. Haas, M. Nagarajan, C. Robson, and N. Sahoo. Artist ranking through analysis of on-line community comments. In *Proc. of the 17th International World Wide Web Conference*, 2008.

[13] A. Heathcote, S. Brown, and D. J. K. Mewhort. The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, 7(2):185–207, 2000.

[14] Y. Hu and M. Ogihara. Nextone player: A music recommendation system based on user behavior. In *Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011*, pages 103–108, 2011.

[15] M. Interiano, K. Kazemi, L. Wang, J. Yang, Z. Yu, and N. L. Komarova. Musical trends and predictability of success in contemporary songs in and out of the top charts. *Royal Society Open Science*, 5(5):171274, 2018.

[16] Y. Kim, B. Suh, and K. Lee. #nowplaying the future Billboard: Mining music listening behaviors of Twitter users for hit song prediction. In *SoMeRA@SIGIR*, pages 51–56, 2014.

[17] N. Koenigstein and Y. Shavitt. Song ranking based on piracy in peer-to-peer networks. In *Proc. of ISMIR 2009*, pages 633–638, 2009.

[18] Christos Koutlis, Manos Schinas, Vasiliki Gkatziaki, Symeon Papadopoulos, and Yiannis Kompatsiaris. T-rec song recognition dataset. `https://doi.org/10.5281/zenodo.3255311`, June 2019.

[19] N. Leibowitz, B. Baum, G. Enden, and A. Karniel. The exponential learning equation as a function of successful trials results in sigmoid performance. *Journal of Mathematical Psychology*, 54(3):338–340, 2010.

[20] G. R. Loftus. Evaluating forgetting curves. *Journal of Experimental Psychology*, 11(2):397–406, 1985.

[21] E. Manoppo, D. P. E. Saerang, and L. Lambey. A comparative study of males and females buyingbehavior toward music playing in store in manado. *Jurnal Berkala Ilmiah Efisiensi*, 16(3):249–257, 2016.

[22] I. L. Menajang. Comparative study of consumer buying behavior at music playing and non music playingshoe stores. *Jurnal EMBA*, 2(3):1321–1328, 2014.

[23] C. Mesnage, R. Santos-Rodriguez, M. McVicar, and T. De Bie. Trend extraction on Twitter time series for music discovery. In *Workshop on Machine Learning for Music Discovery, 32nd International Conference on Machine Learning*, 2015.

[24] J. M. J. Murre and J. Dros. Replication and analysis of Ebbinghaus' forgetting curve. *PLOS ONE*, 10(7):e0120644, 2015.

[25] J. Ren, J. Shen, and R. J. Kauffman. What makes a music track popular in online social networks? In *Proc. of the 25th International Conference Companion on World Wide Web*, pages 95–96, 2016.

[26] F. E. Ritter and L. J. Schooler. The learning curve. In *International encyclopedia of the social and behavioral sciences*, pages 8602–8605. Elsevier Science Ltd, Oxford, 2001.

[27] D. Rohrer, K. Taylor, H. Pashler, J. T. Wixted, and N. J. Cepeda. The effect of overlearning on long-term retention. *Applied Cognitive Psychology*, 19(3):361–374, 2005.

[28] M. Schedl. Analyzing the potential of microblogs for spatio-temporal popularity estimation of music artists. In *Proc. of IJCAI 2011*, 2011.

[29] M. Schedl, T. Pohle, N. Koenigstein, and P. Knees. What's hot? Estimating country-specific artist popularity. In *Proc. of ISMIR 2010*, pages 117–122, 2010.

[30] N. J. Slamecka and L. T. Katsaiti. Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology*, 14(4):716–727, 1988.

[31] R. K. Srivastava. Impact of musical fits and image of different malls on consumer purchase behaviour. *International Journal of Indian Culture and Business Management*, 12(3):318–341, 2016.

[32] I. Mozetic V. Cerqueira, L. Torgo. Evaluating time series forecasting models: An empirical study on performance estimation methods. *arXiv:1905.11744*, 2019.

[33] D. Wang, C. Song, and A. L. Barabási. Quantifying long-term scientific impact. *Science*, 342(6154):127–132, 2013.

[34] J. T. Wixted. Analyzing the empirical course of forgetting. *Journal of Experimental Psychology*, 16(5):927–935, 1990.

[35] R. Yalch and E. Spangenberg. Effects of store music on shopping behavior. *The Journal of Consumer Marketing*, 7(2):55–63, 1990.

[36] L. C. Yang, S. Y. Chou, J. Y. Liu, Y. H. Yang, and Y. A. Chen. Revisiting the problem of audio-based hit song prediction using convolutional neural networks. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 621–625, 2017.