This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MIS.2020.3026000, IEEE Intelligent Systems

An Emotional Recommender System for music

Vincenzo Moscato, Antonio Picariello and Giancarlo Sperlí

Abstract—Nowadays, Recommender Systems have become essential to users for finding "what they need" within large collections of items. Meanwhile, recent studies have demonstrated as *user personality* can effectively provide a more valuable information to significantly improve recommenders' performance, especially considering behavioural data captured from social network logs. In this work, we describe a novel music recommendation technique based on the identification of personality traits, moods and emotions of a single user, starting from solid psychological observations recognized by the analysis of user behavior within a social environment. In particular, users personality and mood have been embedded within a content-based filtering approach to obtain more accurate and dynamic results. Several experiments are then reported to show effectiveness of user personality and mood recognition recommendation, thus encouraging research in this direction.

Index Terms—Recommender Systems, User Personality, Multimedia.

1 INTRODUCTION

When we search for and listen to music, we have a need that is often connected to our specific emotional status: when we are sad, for example, we prefer a piece of classical music by Brahms or Beethoven, or look for sensations linked to memories in a pop sound. The relationships between listening to music and emotions has long been studied by philosophers, psychologists and anthropologists: in all of these application domains, however, a computational approach is necessary that allows both the general mood of a user and the emotional aspect to be identified at a specific time of day or in a period of everyone's life.

In this work, we propose the design and implementation of a music recommender system based on the identification of personality traits, moods and emotions of a single user, starting from solid psychological observations recognized by the analysis of user behavior within a social environment.

Generally speaking, *Recommender Systems* have nowadays become an essential component for intelligent browsing of large collections of objects, thus supporting users to discover "what they need" within the more and more wide ocean of digital information. In our everyday life, recommenders can suggest which movie or photo to watch (e.g., in Youtube and Flickr/Instagram), which kind of music to listen or items to buy (e.g., in Spotify and Amazon), who can invite to a social network (e.g., in Facebook) what travels to do or restaurant to try (e.g., in Booking or TripAdvisor).

More formally, a recommendation process may be modelled as a ranking problem, the objective being to *rank* several *items* from a given set $O = \{o_1, \ldots, o_n\}$. More in details, a recommendation algorithm computes for each *user* u_i belonging to a given community $U = \{u_1 \ldots, u_m\}$ and for each item $o_j \in O$, a score/utility $r_{i,j}$ representing the "interest" of a user u_i into an item o_j.

In *Content-Based Filtering* (CBF) [1], $r_{i,j}$ is estimated exploiting *ratings* assigned to "similar" items. Limitations of this approach are mainly related to the fact that recommendation does not take advantage from behaviour of other users. CBF methods strongly depend on the used similarity function and may suggest only items that are similar to those already analyzed by the same users.

Differently, *Collaborative Filtering* (CF) methods [1] calculate $r_{i,j}$ considering the opinions of other users: the item utility for a given user is predicted on the basis of ratings assigned to the same items by "similar" users. CF suffer of the famous *cold start problem*: a recommender is unable to generate recommendations for new users, lacking initial scores.

To overcome the discussed drawbacks, CBF and CF techniques are generally combined within *hybrid* approaches using different strategies [1].

Furthermore, a novel requirement for a modern recommendation strategy is the use of *context*, e.g., location, observed objects, environmental conditions, and so on, and sometime tools based on the context are called *Context Aware Recommender Systems* [1]: in particular *pre-filtering* data on the basis of users' preferences and needs and *post-filtering* the results on the basis of context may significantly improve the performances, reducing the amount of items to rank and/or the number of suggested objects, respectively.

Recommendation engines have had a large application also in exploring multimedia collections [2]. Audio and music recommender systems gained a great attention due to the diffusion of streaming services as Spotify or Amazon Music platforms, making great use of the information users' profile and behavior.

Nevertheless, several challenges have to be analyzed about user modeling to support different analytics as shown in [3]. In this sub-domain, the related approaches mainly model and exploit users behaviour history considering the ratings provided by users within a wide community, such as an *Online Social Networks* (OSN). Of course, such methods do not perform well with respect to personal factors and in presence of small behaviour history.

Meanwhile, recent studies have demonstrated as user

• *V.Moscato, A. Picariello and G.Sperlí are with the Department of Electrical and Information Technology, University of Naples "Federico II", Via Claudio 21, 80125, Italy.*
E-mail: {vincenzo.moscato, antonio.picariello giancarlo.sperli}@unina.it

Manuscript received xxxx.

personality can effectively provide a more valuable information to build personalized recommender systems in which *diversity*, *popularity*, and *serendipity* of recommendations are the first goals, taking into account the behavioural data captured from OSNs logs.

In psychology, *personality* is considered as the main source for different user preferences and behaviour: there exist, in addition, straight correlations between personality and user's model in recommender systems.

One of the most diffused model that permits to quantitatively measure user personality and can be very helpful within recommender systems is the *Big Five Model*: Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism – also known as OCEAN – are the main user personality traits. The measurement of the personality factors can be obtained using questionnaires with text mining algorithms on social media streams. In particular, "background" (e.g. education, date and place of birth, etc.) and "demographic" data (e.g., place of residence, age, gender, family status, etc.) are often exploited for identifying personality traits together with user emotional states.

Several techniques have been proposed in the literature for automatic personality prediction. They can be classified over two different dimensions: i) *personality recognition* (inferring self-assessed personality traits) and ii) *personality perception* (analyzing the way in which the personality of a given user is perceived). Concerning the first issue, different machine learning approaches have been investigated to predict user personality traits on the basis of users social behaviours. Tadesse et al. [4] compare four machine learning models on *myPersonality* dataset and perform the correlation between each feature set and personality traits.

In addition, various recent papers have shown that personality was successful at facing the cold-start problem, making group recommendations, addressing cross-domain preferences and at generating diverse recommendations [5] using as main strategy a dynamic personality-based greedy re-ranking approach.

In this paper we describe a music recommender system in which we exploit users personality traits as a general model, and provide recommendations based on general mood and emotion. In particular, we combine users personality with their behaviour, preferences, and needs within a content-based recommendation strategy, and also investigate different ways to infer users' personality traits from user-generated data of social networking sites.

The paper is organized as in the following. Section 2 discusses the related work about music recommender systems, focusing also on Content-Based Music Information Retrieval (CB-MIR). Section 3 describes the proposed recommendation framework, while Section 4 presents the related experimentation. Section 5 eventually discusses conclusions and some future work.

2 Related Works

The large amount of digital tracks available online have brought new challenges for recommender systems that need proper methodologies and strategies in order to extract useful information both from users and contents that can be exploited within the recommendation process.

One the first proposal is by Cheng et al. [6]: they proposed a venue-recommender system (*VenueMusic*), relying on a location-aware topic model, adopted for identifying suitable songs for different venues. Furthermore, a tag-aware dynamic music recommendation framework has been developed by Zheng et al. [7], combining tag space and user interaction information. Form another hand, Chen et al. [8] investigated how social influence can improve the performance of a recommender system on the basis of meta path-based similarity measure. A collaborative filtering method, which computes similarity between two users on the basis of playing coefficients, has been then proposed in [9] for music recommendation.

Nevertheless, these recommendation systems are mainly based on user preferences or past listened songs without considering user's mood. In fact, emotion recognition is becoming an hot topic in *Content-Based Music Information Retrieval* (see [10] for more details).

In according to such a trend, Ayata et al. [11] developed a methodology to learn user's emotion by signals obtained from physiological sensors, that can be integrated in any collaborative or content based recommendation framework. In [12], the authors combine content and artists mood similarity based filtering in a unified recommendation framework, called *MoodPlay*. Furthermore, Shen et al. [13] developed a Personality and Emotion Integrated Attentive model (PEIA) combining three different types of features: i) Personality oriented user features (i.e. demographics, textual and social behavior); ii) Emotion-Oriented User Feature (i.e. temporal and emotion) and iii) music features (i.e. metadata, acoustic and emotion).

The majority of the literature approaches defines personality mainly on the basis of textual features (user's post and music's lyrics) without considering the related intrinsic mood which is often the main motivation for the user's choices.

In this paper, we propose a music recommender system based on the identification of personality traits, moods and emotions of a single user. In particular, we considered jointly user's personality and mood for recommending items to final user with respect to [14], which relies only on mel-spectogram's analysis to infer user's personality. In addition, we also embedded mood in a content based recommender with respect to [13], which is mainly based on textual analysis for computing user personality.

3 The Recommendation Framework

Figure 1 shows at a glance the proposed recommendation process for audio contents' suggestion. It works following three different steps:

a *User personality recognition*: user personality is computed in terms of Big Five components, considering user's behavior within social networks.
b *Mood detection*: the last accessed objects are analyzed to discover current user mood, which is then used to refine our recommending strategy.
c *Content-based recommendation*: audio contents are suggested to users on the basis of their content and the related similarity.

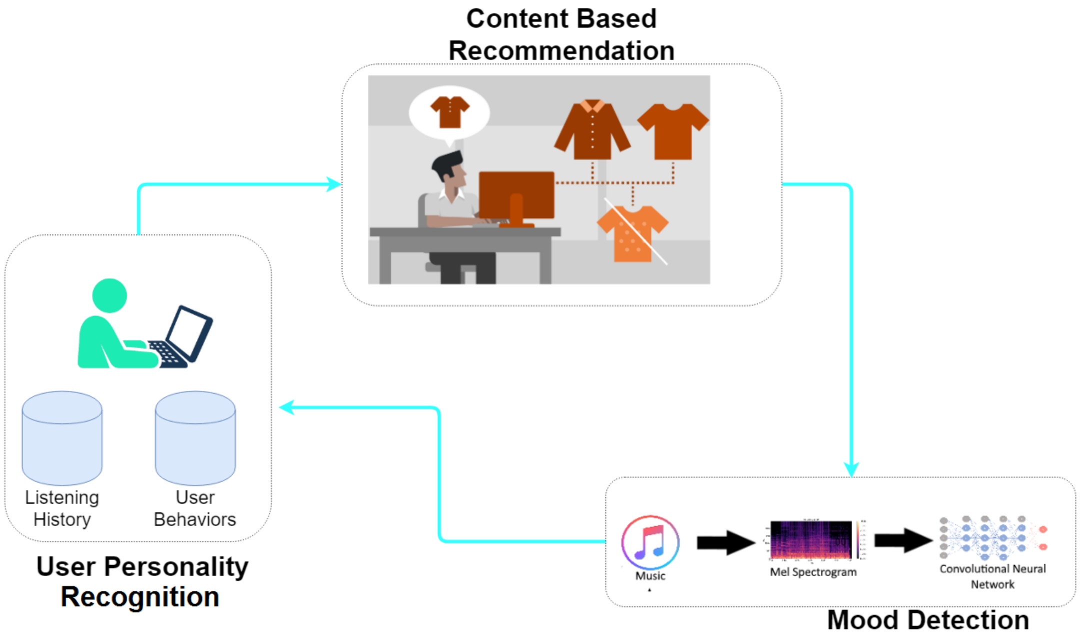

Fig. 1: Recommendation Process Workflow.

3.1 User Personality Recognition

Our approach for personality recognition is summarized in Figure 2. Personality recognition's main goal is to automatically infer users' personality considering their actions/behaviors extracted from OSNs.

As previously discussed, in the system we take into account several types of information that psychologists' thinks may affect user personality: *Stable* – user background – *Semi-Stable* – demographic data – *Not-Stable*, the user mood.

Figure 2 describes in details the sub-system workflow. In particular we notice that not-stable information may be derived by means of the user published images, user status updates and, in specific OSNs, the number of "likes".

In more detail, the user's status updates are processed through a NLP pipeline: a tagger based on dictionaries has been implemented for handling emoticons and emojis, that sometimes represent the basic emotional features. The tagged status is then analyzed for computing i) sentiment scores and ii) psychological features; in particular, we have used the *Sentiwordnet*[1] and *Emoji Sentiment Ranking*[2] for computing the sentiment score, and the *MRC Psycholinguistic Database* [3] (a database of psycholinguistic information regarding 26 different psycholinguistic properties) for the psychological features.

Eventually, sentiment and psychological information are then combined with the number of "likes", if present, and the stable and semi-stable information: in this way, we train five different classifiers, one for each OCEAN personality trait.

In addition, we have extracted personality information from the images the user has posted on the social networks, by means of *Convolutional Neural Networks* (CNNs) (again, one for each personality trait), as proposed by [15].

The final Five Personality Traits are obtained combining the results of the two separated sets of classifiers. We adopted a simple linear combination as in the following:

$$\left(\frac{n_{status}}{n_{status}+n_{photos}}\right)\cdot F_{ML}+\left(\frac{n_{photos}}{n_{status}+n_{photos}}\right)\cdot F_{DL} \quad (1)$$

F_{ML} and F_{DL} being respectively the machine learning and deep learning prediction value with respect to a personality trait F, while n_{status} and n_{photos} being the number of user's status updates and pictures published by the user.

The obtained OCEAN user personality traits are subsequently mapped into Mehrabian's *Pleasure-Arousal-Dominance* (or PAD) emotional state space [16] following rules in Table 1.

P =	.59 · Agree + .19 · Neur + .21 · Extra
A =	-.57 · Neur + .30 · Agree + .15 · Open
D =	.60 · Extra - .32 · Agree + .25 · Open + .17 · Conc

TABLE 1: Relationships between the Big-five and PAD scales.

We have mapped both items and users into the PAD emotional space, except for the Dominance dimension that has not been considered [16].

3.2 Mood detection and Recommendation Updating

This step has the goal to update user's mood computing audio's *Arousal* and *Pleasure*, sometimes also called *Valence*, values based on the listened songs in a given time interval.

The actual mood state, which initially corresponds to user's personality, is then continuously updated on the basis

1. https://sentiwordnet.isti.cnr.it
2. http://kt.ijs.si/data/Emoji_sentiment_ranking/
3. http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm

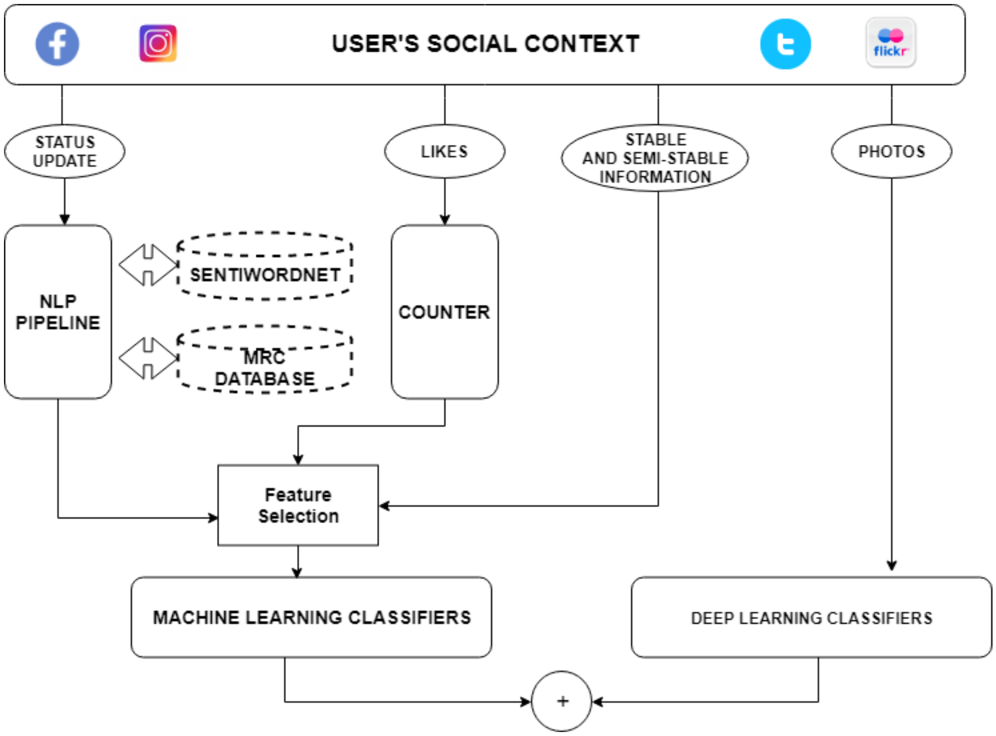

Fig. 2: User Personality Recognition.

of the previous mood and emotional states related to latest heard items (see [17] for more details) using the following formulas:

$$M_{\text{t}} = w_1 \cdot M_{\text{t-1}} + w_2 \cdot \phi(e) \tag{2}$$

$$\phi(e) = \frac{e_{\text{t}} + e_{\text{t-1}} + e_{\text{t-2}} + ... + e_{\text{t-n}}}{n} \tag{3}$$

where t represents the time scale, $M_{\text{t-1}}$ is the previous mood, ϕ(e) is the history of previous emotion of songs, defined according to equation 3, and w_1, w_2 are particular weights. The values to be assigned as coefficients can be measured based on individuals possible mood swings.

Summing up, we update and predict PAD mood traits on the basis of the information regarding personality traits, estimated emotion states and mood history probability distribution.

We have computed song's Arousal and Valence values using proper CNNs. Due to the fact that the time-frequency representation of a given audio contains a large amount of information useful to identify similar sounds, we used the *Mel-Spectrogram* time-frequency model.

It can be seen as a particular function, namely $S_{mel}(f,t)$, where f represents the frequencies' center of the Mel filters and t is the time frame of the *Short-Time Fourier Transform* (STFT).

For each song listened by a user, we extracted the Mel-Spectrogram that becomes the input for a VGG CNN.

Indeed, we have used two VGG networks, one for Arousal and one for Valence, composed by five consecutive 2-dimensional convolution layers with different feature maps of size 3, batch normalization and 3 fully connected layers as output of the network.

3.3 Content-based Recommendation

Audio objects are mapped into specific points of the PAD space following the strategy described in the previous subsection 3.2, using song's Mel-Spectogram to compute arousal and valence values.

Considering the large number of audio object, we organize them using a *Ball-Tree* data structure, considering Pleasure and Arousal values as the main dimensions. Items are then suggested to users on the basis of a minimum Euclidean distance criterion with respect to the mood of the examined user mood, performing the Nearest Neighbor Query search on the Ball-Tree.

4 Experiments

4.1 The Experimental Protocol

The experimental protocol has been conducted following three different steps:

- We first evaluated the accuracy of our user personality recognition module comparing it with other approaches in the literature, with respect to two different standard datasets;
- We then tuned our mood detection module using different CNNs in terms of *Mean Squared Error* (MSE);
- Successively, we compared our classification approach using hyper-parameters identified in the previous stage with respect to other ones (SVM,CNN Deezer) in terms of R^2;
- Eventually, we compared the proposed recommendation strategy with respect to two well-known rating-based techniques, namely UPCC [18] and IPCC [18] using several recall based metrics.

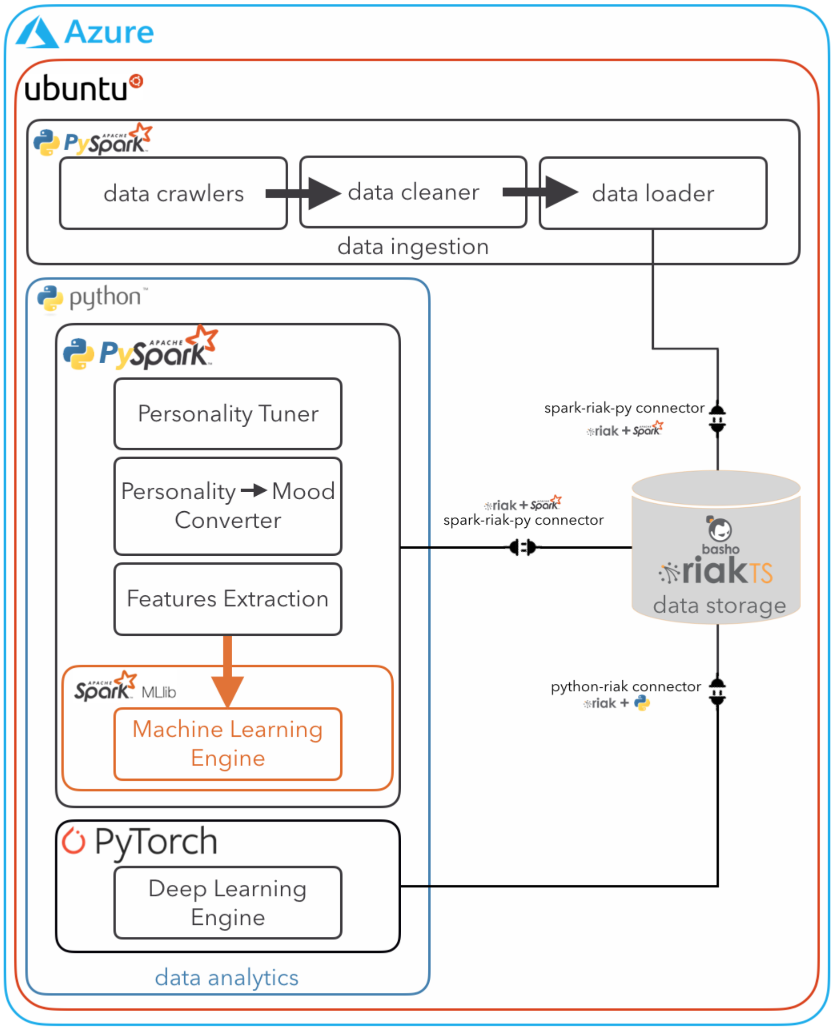

Fig. 3: Architecture of User Personality Recognition module.

4.2 Datasets and Computing Environment

Concerning mood detection and recommendation, the experiments have been performed on the Deezer dataset[4], composed by 18.645 songs with several metadata (i.e. deezer song id, artist name, track name, valence, arousal and so on).

In particular, we trained the Mood Detection Classifiers (i.e, CNNs) on approximately 60% of the original dataset using a ten-fold cross validation. The residual 40% of the dataset is, then, divided into two equal parts to respectively validate and test each trained model.

We randomly split the Deezer dataset respecting the constraint that each artist belongs only to a single set to avoid any correlations between artist and moods, according to the same experimental protocol described in [14]. These tree subsets (training, validation and test) are respectively composed by 11.267, 3.863 and 3.514 songs.

In turn, for evaluating effectiveness of user personality recognition, we used two different datasets: *myPersonality*[5] and *PsychoFlickr*[6]. The first one is composed by 4.3 millions of users, 22 millions of status updates, 19 millions of likes and 3.1 millions of Big 5 scores obtained filling a personality questionnaire; the second one is constituted by 6000 favourite pictures for 300 Flickr users, to which are assigned the Big 5 scores using assessed traits based on the BFI-10 questionnaire (filled by 10 accessors that looked the latest 200 favourite user pictures).

Concerning implementation details of this component (see Figure 3), data ingestion has been realized using proper API, data storage has been obtained through the NoSQL database Basho Riak TS, data analytics for user personality recognition has been performed exploiting Apache PySpark and Pytorch facilities.

The evaluation has been carried out using Google Colab platform [7] equipped with:

- GPU: 1xTesla K80 , compute 3.7, having 2496 CUDA cores , 12GB GDDR5 VRAM;
- CPU: 1 x single core hyper threaded Xeon Processors @2.3Ghz;
- RAM: 12.6 GB Available;
- Disk: 33 GB Available.

4. https://github.com/deezer/deezer_mood_detection_dataset
5. http://mypersonality.org/
6. http://vips.sci.univr.it/dataset/psychoflickr/PsychoFlickr.rar
7. https://colab.research.google.com/

4.3 User Personality Recognition Accuracy

We decided to perform two sets of experiments: the first one on myPersonality dataset, in order to evaluate the accuracy of the adopted machine learning classifiers, and the second one on the PsychoFlickr dataset, for testing performances of the deep learning classifiers.

As already discussed, we split the user personality recognition problems into five distinct binary classification subproblems (one for each trait). Furthermore, we partitioned each trait into *low*, *medium* and *high* classes, composed by the values below the first quartile, between the second and the third quartile, and above the third quartile, respectively. For each trait we have chosen only the values belonging to the high or low groups in order to guarantee greater separation between the two classes.

Table 2 summarizes the number of users for each class in both datasets.

It is easy to note that PsychoFlickr (E_2) is a balanced dataset, whilst we performed a down sampling with an approximate ratio of 1.5 between the two considered classes to reduce the displacement of myPersonality dataset (E_1).

Missing values have been managed through a mixed approach in which for some fields (such as *gender*, *relationship_status*, *interested_in*, *locale*) we replace them with the ND label because we considered the lack of a value as the user's willingness to not disclose this personal information. In turn, for other fields (such as *network_size*, *tot_likes* and *age*) single values are replaced with the related mean.

Eventually, since the features extracted from our framework are too many (i.e., 47 different features), we decided to adopt a *Forward Features Selection* strategy evaluated with a *10-fold Cross Validation*. In particular, the experiments on *PsychoFlickr* dataset were performed using different classic machine learning systems – such as the Naive-Bayes, Logistic Regression, SVM, Random Forest, GBT – and using deep networks such as Alexnet, Resnet50, Squeeznet, Densenet, Vgg.

4.3.1 Results for MyPersonality dataset

The results are shown in Table 3. It contains the accuracy on the training and the test sets and the number of selected features by our Forward Features Selection algorithm, for each classifiers. Then, the obtained best results are compared with those presented by [15] using a self-assessed and attributed-assessed strategy (see Tables 4 and 5 for the details, respectively).

However, our results mark a significant improvement compared to the literature. Furthermore, this results could be subject to further improvements through a more careful fine-tuning of the classifiers' hyper-parameters.

Beyond the achieved results, it was also interesting to understand the choice of features done by the algorithm. In particular, we noticed that the Conscientiousness factor, to be predicted, requires on average a greater number of features (about 22) while the Neuroticism factor is absolutely the one that requires less features (about 14).

Moreover, by observing the subset of optimal features chosen by each classifier, we realized several interesting aspects.

Although for the Openess factor each classifier chooses a different feature as first selected, it is possible to observe that the features with the most discriminating power are those linked to the negativity of the posted status.

As for the Conscientiousness factor, all the classifiers have chosen as features more discriminating the age and the average number of emojis and emoticons published in the status update.

The Extroversion, on the other hand, was closely linked to the size of the network of friendships on the social network as well as to the number of positive posted states.

The Agreeableness factor is instead the one that shows a greater heterogeneity in the choices made by the classifiers, that chose different features to be included in the optimal subset.

Furthermore, the Neuroticism factor was strongly linked to the frequency of social interactions (number of updates and likes): 3 classifiers out of 5 choose this as most discriminating feature, while the remaining 2 choose it for fourth and second main feature, respectively.

Finally, Table 5 shows the results obtained by the proposed approach using the unbalanced dataset that considers the three classes (low, medium and high). As easy to note in Table 5, reducing classes' separation achieves less accurate results.

4.3.2 Results for PsychoFlickr dataset

The results are summarized in Table 7, showing the accuracy obtained for both the training and the test sets, also exploiting multimedia content and a comparison with the self-assessed strategy in [15]. Furthermore, we show in Table 8 the results of our approach using all classes (low, medium, high), that achieves less precise results, due to the decrease of class separation.

Finally, we compared the results of both experiments E_1 and E_2: we can note that the approach in which, the feature extraction is entrusted to human, leads to better results than the automatic one (see Figure 4).

4.4 Mood Detection Accuracy

We first evaluated the mood detection module varying the chosen CNNs (VGG11,VGG13,VGG16 and RES50) that work on songs' mel-spectogram with 128 mel-filters and a sampling frequency of 8kHz. Audios have been crawled using the Deezer API[8]. We have divided such a task into two distinct regression problems, one for Arousal and another one for Valence (see Figure 5), PAD values being independent [16].

Table 10 analyzes the performances of the proposed approach on the test set for Valence and Arousal measures in terms of Mean Square Error (MSE). RES50 has always worse results and requires a greater number of epochs with respect to other CNNs for both Valence and Arousal.

In addition, our mood detection module, using the best parameters identified in the previous evaluation (a learning rate of 0.0001 and a batch size of 64), has been compared with respect to SVM and Deezer CNN [14], that use respectively Mel-Spectogram and MFCC, according to R^2 measure, whose results are shown in Table 11.

8. https://developers.deezer.com/api

	E_1 - myPersonality					E_2 - PsychoFlickr				
	O	C	E	A	N	O	C	E	A	N
low	9549	8830	15434	10808	9998	113	122	75	110	104
high	57616	45518	41883	52934	46167	119	121	76	125	111

TABLE 2: Distribution of OCEAN values for users within the two datasets.

		Openess	**Conscientiousness**	**Extroversion**	**Agreeableness**	**Neuroticism**
Nave-Bayes	**Train**	0.6142	0.6583	0.7255	0.5969	0.6311
	Test	0.6507	0.6594	0.7325	0.6194	0.6456
	#features	6	10	10	25	4
Logistic Regression	**Train**	0.7187	0.7240	0.7378	0.7706	0.7395
	Test	0.6972	0.7163	0.7438	0.7007	0.6853
	#features	25	20	16	27	18
SVM	**Train**	0.6929	0.7257	0.6979	0.6788	0.6679
	Test	0.6840	0.7059	0.6900	0.6548	0.6346
	#features	17	20	21	21	14
Random Forest	**Train**	0.7905	0.7706	0.7725	0.7543	0.7357
	Test	0.7110	0.7007	0.7507	0.7313	0.7124
	#features	16	30	13	25	22
GBT	**Train**	0.8700	0.7782	0.7324	0.7176	0.7567
	Test	0.6884	0.6853	0.6679	0.6435	0.6976
	#features	21	20	20	16	15

TABLE 3: Accuracy results for dataset E_1.

Our approach - Self Assessed					[15] - Self Assessed				
O	C	E	A	N	O	C	E	A	N
0.71	0.72	0.75	0.73	0.71	0.54	0.55	0.54	0.54	0.54

TABLE 4: Comparison of the results on E_1: self vs. self.

		Openess	**Conscientiousness**	**Extroversion**	**Agreeableness**	**Neuroticism**
Nave-Bayes	**Train**	0.5792	0.6228	0.6933	0.5645	0.6127
	Test	0.6178	0.6166	0.7022	0.5833	0.6201
	#features	8	11	11	25	7
Logistic Regression	**Train**	0.6841	0.6932	0.7012	0.7542	0.7101
	Test	0.6699	0.6956	0.7189	0.6743	0.6599
	#features	27	22	17	27	20
SVM	**Train**	0.6699	0.6912	0.6689	0.6576	0.6322
	Test	0.6599	0.6798	0.6701	0.6289	0.6106
	#features	18	22	23	23	15
Random Forest	**Train**	0.7685	0.7596	0.7601	0.7320	0.7151
	Test	0.6899	0.6801	0.7244	0.7089	0.6744
	#features	18	30	16	28	25
GBT	**Train**	0.8455	0.7476	0.7024	0.6891	0.7122
	Test	0.6543	0.6623	0.6499	0.6123	0.6696
	#features	23	21	21	19	16

TABLE 5: Accuracy results for the dataset E_1 using all classes.

As easy to note in Table 11, the proposed approach achieves comparable results with respect to the state-of-the-art although the training of SVM and CNN Deezer approaches are based on augmentation techniques improving the number of training samples using emotion invariant operations (pitch shifting and lossy encoding).

4.5 Recommendation Effectiveness

Furthermore, we evaluated the overall recommendation approach by asking 50 users, whose age ranges between 22 to 55 years old having different degree levels, to interact with our recommendation system, for evaluating the recommended items.

Users have also completed a psychological test for the evaluation of the related psychological profile in terms of 5 Big Five components that correspond to the basic mood of a given user. In addition, they provide also the authorization to crawl the last 12 heard songs. Each user has rated a dataset of 500 songs and labelled each component item marking it as of interest (100 at most) or not with respect to her/his current mood. The objects of the same dataset were then ranked using our approach, UPCC and IPCC strategies with respect to the human ground truth, that is defined according to the previous constraints, varying the number of suggested items to compute R@.

Table 10 shows the obtained results in terms of recall with respect to different recommendation output sizes. As easy to note in Table 12, our approach obtains better results with respect to IPCC and UPCC that are only focused on item or user respectively exploiting the related static ratings and similarity criteria. In turn, our strategy takes a great advantage of the current user mood, and by which a content-based technique, it can significantly improve the recommendation's performances.

Our approach - Attributed					[15]- Attributed Assessed				
O	C	E	A	N	O	C	E	A	N
0.71	0.72	0.75	0.73	0.71	0.61	0.67	0.65	0.64	0.69

TABLE 6: Comparison of the results on E_1: attributed vs. attributed.

		Openess	Conscientiousness	Extroversion	Agreeableness	Neuroticism
Alexnet	**Train**	0.5805	0.5931	0.5974	0.5813	0.5465
	Test	0.5504	0.5401	0.5491	0.5553	0.5234
Resnet50	**Train**	0.6233	0.5731	0.5698	0.6234	0.6432
	Test	0.6051	0.5668	0.5621	0.5973	0.6123
Squeezenet	**Train**	0.6051	0.5432	0.5876	0.5575	0.5567
	Test	0.5960	0.5411	0.5323	0.5438	0.5365
Densenet	**Train**	0.6009	0.5991	0.5787	0.5643	0.5434
	Test	0.5617	0.5512	0.5384	0.5243	0.5564
Vgg	**Train**	0.6322	0.6102	0.5543	0.5776	0.5576
	Test	0.6022	0.5898	0.5221	0.5349	0.5234

TABLE 7: Accuracy results for dataset E_2.

		Openess	Conscientiousness	Extroversion	Agreeableness	Neuroticism
Alexnet	**Train**	0.5596	0.5712	0.5615	0.5499	0.5122
	Test	0.5324	0.5299	0.5287	0.5304	0.5099
Resnet50	**Train**	0.6043	0.5499	0.5408	0.6041	0.6199
	Test	0.5812	0.5427	0.5399	0.5699	0.5954
Squeezenet	**Train**	0.5823	0.5201	0.5598	0.5321	0.5301
	Test	0.5669	0.5125	0.5101	0.5198	0.5171
Densenet	**Train**	0.5797	0.5657	0.5583	0.5499	0.5212
	Test	0.5525	0.5379	0.5214	0.5087	0.5303
Vgg	**Train**	0.6101	0.5953	0.5298	0.5503	0.5365
	Test	0.5834	0.5676	0.5067	0.5138	0.5098

TABLE 8: Accuracy results for the dataset E_2 using all classes.

Our approach - Self Assessed					[15] - Self Assessed				
O	C	E	A	N	O	C	E	A	N
0.61	0.59	0.56	0.60	0.61	0.54	0.55	0.54	0.54	0.54

TABLE 9: Comparison of the results on E_2 : self vs. self.

5 Conclusions

A personality-based recommender system has been discussed and analyzed, using the Big Five psychological model and considering both user profile and mood for a content-based strategy. As shown by the obtained experimental results, our work takes an enormous advantage of these two kinds of information to support in a very effective manner browsing of audio collections with respect to classical rating-based recommendation approaches.

Future work will be devoted to improve our strategy also in presence of a poor musical chronology. In addition, we plan to extend our experiments using different OSNs to better capture user personalty and to compare our results with commercial techniques as those exploited by Spotify and Amazon music.

References

[1] F. Ricci, L. Rokach, and B. Shapira, "Recommender systems: introduction and challenges," in *Recommender systems handbook*. Springer, 2015, pp. 1–34.

[2] M. Albanese, A. dAcierno, V. Moscato, F. Persia, and A. Picariello, "A multimedia recommender system," *ACM Transactions on Internet Technology (TOIT)*, vol. 13, no. 1, p. 3, 2013.

[3] P. De Bra, "Challenges in user modeling and personalization," *IEEE Intelligent Systems*, vol. 32, no. 5, pp. 76–80, 2017.

[4] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Personality predictions based on user behavior on the facebook social media platform," *IEEE Access*, vol. 6, pp. 61 959–61 969, 2018.

[5] R. P. Karumur, T. T. Nguyen, and J. A. Konstan, "Personality, user preferences and behavior in recommender systems," *Information Systems Frontiers*, vol. 20, no. 6, pp. 1241–1265, 2018.

[6] Z. Cheng and J. Shen, "On effective location-aware music recommendation," *ACM Trans. Inf. Syst.*, vol. 34, no. 2, Apr. 2016. [Online]. Available: https://doi.org/10.1145/2846092

[7] E. Zheng, G. Y. Kondo, S. Zilora, and Q. Yu, "Tag-aware dynamic music recommendation," *Expert Systems with Applications*, vol. 106, pp. 244 – 251, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417418302446

[8] J. Chen, P. Ying, and M. Zou, "Improving music recommendation by incorporating social influence," *Multimedia Tools and Applications*, vol. 78, no. 3, pp. 2667–2687, 2019.

[9] D. Snchez-Moreno, A. B. Gil Gonzlez, M. D. Muoz Vicente, V. F. Lpez Batista, and M. N. Moreno Garca, "A collaborative filtering method for music recommendation using playing coefficients for artists and users," *Expert Systems with Applications*, vol. 66, pp. 234 – 244, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417416304973

[10] Y. V. S. Murthy and S. G. Koolagudi, "Content-based music information retrieval (cb-mir) and its applications toward the music industry: A review," *ACM Comput. Surv.*, vol. 51, no. 3, Jun. 2018. [Online]. Available: https://doi.org/10.1145/3177849

[11] D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion based music recommendation system using wearable physiological sensors," *IEEE Transactions on Consumer Electronics*, vol. 64, no. 2, pp. 196–203, 2018.

[12] I. Andjelkovic, D. Parra, and J. ODonovan, "Moodplay: Interactive music recommendation based on artists mood similarity," *International Journal of Human-Computer Studies*, vol. 121, pp. 142 – 159, 2019, advances in Computer-Human Interaction for Recommender Systems. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1071581918301654

[13] T. Shen, J. Jia, Y. Li, Y. Ma, Y. Bu, H. Wang, B. Chen, T.-S. Chua, and W. Hall, "Peia: Personality and emotion integrated attentive model for music recommendation on social media platforms," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, 2020, pp. 206–213.

[14] J. R.-L. M. M. Remi Delbouys Romain Hennequin Francesco Piccoli, "Music mood detection based on audio and lyrics with deep

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MIS.2020.3026000, IEEE Intelligent Systems

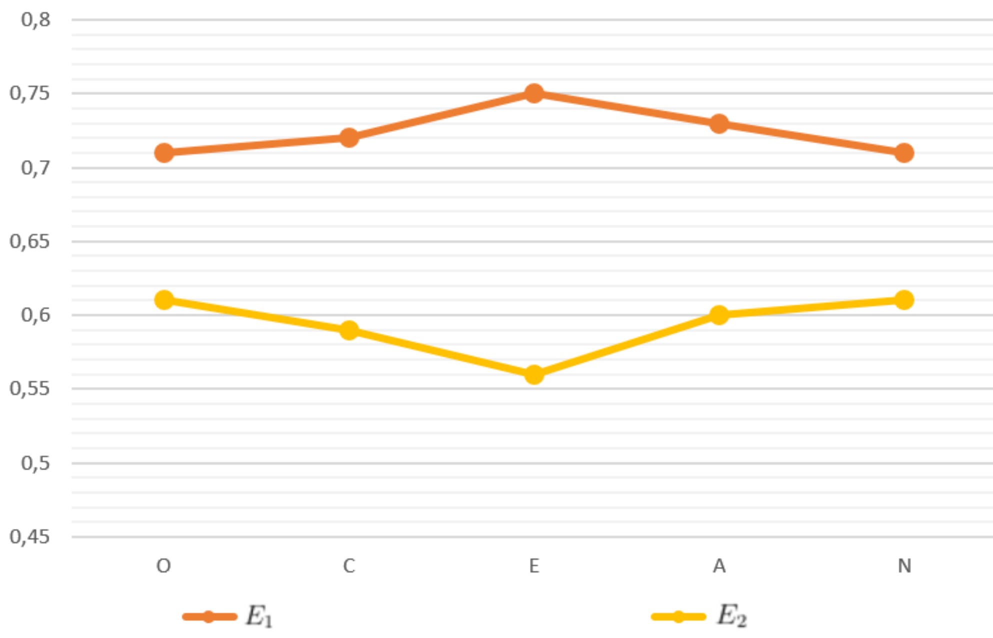

Fig. 4: Comparison of the results of E_1 and E_2.

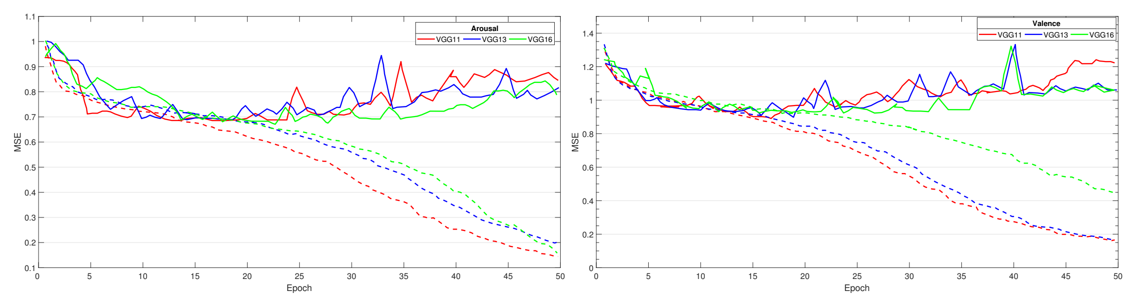

Fig. 5: VGG11 (red),VGG13 (blu),VGG16 (green) Arousal and Valence MSE. The dashed line represents the results on the train set and the continuous line represents the results on the validation.

CNN	Arousal		Valence	
	Epoch	MSE	Epoch	MSE
VGG11	13	0.692	17	0.8905
VGG13	19	0.6916	19	0.9026
VGG16	23	0.6868	21	0.9360
RES50	93	0.7109	90	0.9718

TABLE 10: CNNs' results on test set
for Arousal and Valence.

Model	Input	Arousal	Valence
SVM	MFCC	0.197	0.117
CNN Deezer	MEL	0.235	0.179
Our approach	MEL	0.23	0.171

TABLE 11: R^2 scores of the different approaches.

	IPCC	UPCC	Our approach
R@50	0.22	0.26	0.63
R@70	0.46	0.52	0.94
R@85	0.52	0.56	1
R@100	0.58	0.6	1

TABLE 12: Recommendation Recall.

neural net," *19th International Society for Music Information Retrieval Conference*, 2018.

[15] C. Segalin, F. Celli, L. Polonio, M. Kosinski, D. Stillwell, N. Sebe, M. Cristani, and B. Lepri, "What your facebook profile picture reveals about your personality," in *Proceedings of the 25th ACM International Conference on Multimedia*, ser. MM '17. New York, NY, USA: ACM, 2017, pp. 460–468. [Online]. Available: http://doi.acm.org/10.1145/3123266.3123331

[16] A. E. Krause and A. C. North, "Pleasure, arousal, dominance, and judgments about music in everyday life," *Psychology of Music*, vol. 45, no. 3, pp. 355–374, 2017.

[17] M. Mehraei and N. I. Akcay, "Pleasure, arousal, and dominance mood traits prediction using time series methods," *IAFOR Journal of Psychology & the Behavioral Sciences*, 2017.

[18] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in artificial intelligence*, vol. 2009, 2009.

Vincenzo Moscato Vincenzo Moscato is an Associate Professor at the Electrical Engineering and Information Technology Department of University of Naples "Federico II".He is member of the Consorzio Interuniversitario Nazionale per lInformatica (CINI), Centro di competenza Regionale per lICT (CeRICT) and International Association of Pattern Recognition, Italian Chapter. His current research interests concern Big Data Infrastructures, Big Data Analytics, Multimedia, Recommender Systems and Social Network Analysis. He was involved in many national and international projects and coordinated some of the them. He has been member and chair of different organizing and program committees of International Conferences, and he has been reviewer and in the editorial boards for many international journals. He is an author of about 150 publications on international journals, conference proceedings and book chapters.

Antonio Picariello is a Full Professor at Department of Electrical Engineering and Information Technology, University of Naples Federico II. He got a ph. d. in Computer Engineering at the University of Napoli Federico II. He is the director of the National Lab of Computer Science, Telematics and Multimedia (ITEM) of the Italian Consortium on Computer Science and Engineering (CINI). He works in the field of Multimedia Database and Multimedia Information Systems, Multimedia Ontology and Semantic Web, Natural Language Processing, Big Data, Big Data analytics and Social Networks Analysis.

Giancarlo Sperlí is a Researcher at the Consorzio Interuniversitario per l'Informatica (CINI). He hold a Masters Degree and a Bachelors Degree in Computer Science and Engineering, both from the University of Naples Federico II and in 2018 he received the Ph.D degree in Information Technology and Electrical Engineering of University of Naples "Federico II". His main research interests are in the area of Cybersecurity, Semantic Analysis of Multimedia Data and Social Networks Analysis.