Towards Emotion-aware Recommender Systems: an Affective Coherence Model based on Emotion-driven Behaviors

Article *in* Expert Systems with Applications · December 2020

DOI: 10.1016/j.eswa.2020.114382

4 authors:

Marco Polignano
Università degli Studi di Bari Aldo Moro

Fedelucio Narducci
Politecnico di Bari

Marco de Gemmis
Università degli Studi di Bari Aldo Moro

Giovanni Semeraro
Università degli Studi di Bari Aldo Moro

Towards Emotion-aware Recommender Systems: an Affective Coherence Model based on Emotion-driven Behaviors

Marco Polignano[a], Fedelucio Narducci[b], Marco de Gemmis[a,*], Giovanni Semeraro[a]

[a] *Department of Computer Science, University of Bari Aldo Moro Via E. Orabona 4, I-70125 Bari, Italy*
[b] *Department of Electrical and Information Engineering, Politecnico di Bari Via E. Orabona 4, I-70125 Bari, Italy*

Abstract

Decision making is the cognitive process of identifying and choosing alternatives based on preferences, beliefs, and degree of importance given by the decision maker to objects or actions. For instance, choosing which movie to watch is a simple, small-sized decision-making process. Recommender systems help people to make this kind of choices, usually by computing a short list of suggestions that reduces the space of possible options. These systems are strongly based on the knowledge of user preferences but, in order to fully support people, they should be grounded on a holistic view of the user behavior, that includes also how emotions, mood, and personality traits influence her choosing patterns.

In this work, we investigate how to include *emotional* aspects in the recommendation process. We suggest that the affective state of the user, defined by a set of emotions (e.g., joy, surprise), constitutes part of choosing situation that should be taken into account when modeling user preferences.

The main contribution of the paper is a general emotion-aware computational model based on affective user profiles in which each preference, such as a 5-star rating on a movie, is associated with the affective state felt by the user

*Corresponding author

Email addresses: marco.polignano@uniba.it (Marco Polignano), fedelucio.narducci@poliba.it (Fedelucio Narducci), marco.degemmis@uniba.it (Marco de Gemmis), giovanni.semeraro@uniba.it (Giovanni Semeraro)

at the time when that preference was collected.

The model estimates whether an unseen item is suitable for the current affective state of the user, by computing an *affective coherence score* that takes into account both the affective user profile and not-affective item features. The approach has been implemented into an Emotion-aware Music Recommender System, whose effectiveness has been assessed by performing in-vitro experiments on two benchmark datasets. The main outcome is that our system showed improved accuracy of recommendations compared to baselines which include no affective information in the recommendation model.
Keywords: Recommender Systems, Affective Computing, Music Recommendation, Decision Making, myPersonality, Emotions

---

1. Introduction

The great availability of personalized services over the Internet has made users more inclined to provide their data in order to obtain accurate suggestions about products to buy, news to read, music to listen to. Today, companies such as Netflix or Amazon collect millions of records about habits of their customers and use them to target their suggestions by exploiting recommender systems (RecSys), that are filtering tools that guide the user in a personalized way to interesting or useful items in a large space of possible options (Burke, 2002; de Gemmis, Lops, and Polignano, 2017). These systems collect information about user preferences, either explicitly by asking users to provide ratings on items, or implicitly by analyzing their actions on items (e.g., downloads, prints, views). Collected data are exploited to build a model of user preferences (user profile), or to discover users having similar interests, with the aim of finding novel items that might be interesting to them. Today, personalization systems could take advantage of digital footprints, i.e. the entire collection of information generated by a person's online activity, such as online searches, purchased items, or posts on social media. In particular, actions performed on Social Media Sites (SMSs) provide valuable information about user tendencies, styles of life, as well as affective and psychological traits (Ryan and Xenos, 2011; Asur and Huberman, 2010; Correa, Hinsley, and De Zuniga, 2010; Malhotra et al., 2012). It is not surprising that social media footprints, such as liked pages, attended events, shared music, etc., are being more and more exploited by RecSys (Kazai, Yusof, and Clarke, 2016; Ma et al., 2017) to infer personality traits (Back et al., 2010), affective states or tendencies (Alm, Roth, and Sproat, 2005; Mohammad and Kiritchenko, 2015; Polignano et al., 2017c), with the aim of improving accuracy of suggestions. Recently, the European Legislation has strongly limited the use of that data without an explicit agreement by the final user. In particular, the GDPR regulation (Goddard, 2017) has placed restrictions on the exploitation of user data in order to limit illegal use. In our work, we used only third-party anonymized publicly available datasets. In the case the system will be implemented in real-world environments, the privacy issue should be addressed, of course.

Several studies in Psychology described how emotions actively interact with the choosing patterns evoked in the subject during decision processes (Picard et al., 2004; Clore, Schwarz, and Conway, 1994; Bechara, Damasio, and Damasio, 2000; Loewenstein and Lerner, 2003b; Frijda and Mesquita, 1994; Cunningham, 1988) and showed that those areas of our brain related to feelings are highly stimulated during decision-making processes (Bechara, 2003).

Therefore, novel techniques to model user preferences and attitudes, according to a holistic view of personalization, are needed. The main contribution of this work is the investigation of the following open issues:

1. To build an *affective* user profile in which preferences are modeled considering affective information;
2. To include an affective profile into a recommendation process.

As a proof-of-concepts, we implemented an Emotion-aware RecSys and performed an evaluation in the music domain. The results showed improved accuracy of recommendations compared to several baselines, including machine learning methods and content-based filtering methods which exploit no affective information to model user preferences. The rest of the paper is organized as follows: Section 2 provides an analysis of the main relevant work for this research, Section 3 describes our original model for generating emotion-aware recommendations, implemented in EMRES that is described in Section 4. Finally, the experimental evaluation and the results are outlined in Section 5.

2. Related Work

The proposed investigation falls at the intersection between Psychology and Computer Science. In this section, we provide some background and related work that allow to place our research in the appropriate context of both areas. In the analysis of Psychology literature, we focus on the role that affect (emotions, mood, personal feelings) could play in decision-making processes, in order to motivate the design choices behind our affect-aware model. Then, we describe some recent attempts to integrate emotions, personality, and mood into personalization methods, in particular into recommender systems.

2.1. Affective influences in decision making

Events that involve decision-making processes are characterized by two main steps: (i) problem formalization and (ii) evaluation of the alternative possibilities. In the first step, the problem and its possible solutions are conceptualized in the mind of the decision maker. In the second one, each option is evaluated by taking into account different perspectives, goals, and expectations. Klein proposed the Naturalistic Decision Model (NDM) (Klein, 2015; Klein, Calderwood, and Clinton-Cirocco, 2010) as a formalization of the intuitive idea that every subject of a decision-making process evaluates different options according to her goals, causal factors, perceptual cues, then she opportunely decides whether an action can be implemented, rejected, or modified for covering the personal unsatisfied constraints. The NDM formalization points out the important role of comparing past experiences, in the same or in similar contexts, during decision making. The main problem is how to conceptualize the role of emotions and personal feelings in this scenario. The issue has been deeply studied in Psychology (Pfister and Böhm, 1992; Pfister and Böhm, 2008; Loewenstein and Lerner, 2003a; Peters, 2006; Fiori et al., 2013). According to traditional approaches of behavioral decision making, choosing is seen as a rational cognitive process that estimates which of various alternative choices would yield the most positive consequences, which does not necessarily entail emotions. Feelings are considered as external forces influencing an otherwise non-emotional process, according to the *influence-on metaphor*. There are different ways in which emotions enter into decision making. In (Peters, 2006), the author focused on the informational value of affect, that is when decision makers intentionally consult their feelings about an option and use that information to guide the decision process. In that situation, four roles of emotions are identified:

1. Information: affect developed through experience provides information about what to choose and what to avoid by marking decision options and attributes by positive and negative feelings;
2. Spotlight: emotions can focus the decision maker's attention on certain aspects of the problem and may alter which information become salient;
3. Motivator: incidental emotions motivate behavior as people tend to act in order to maintain or attain positive mood states;
4. Currency: affect can provide a common currency for experiences, thus enabling people to compare arguments on a common underlying dimension.

These roles can be found even mixed when selecting an option. For example, when the decision maker takes into account previous experiences, past choices adopted in situations which are similar to the current decision problem can be evaluated according to the positive or negative feelings they evoke (information role). Then, options can be compared by simpler affective evaluations, rather than by attempting to make sense out of a multitude of conflicting logical reasons (currency role). These ideas are endorsed and extended in the work by Pfister and Bohm (Pfister and Böhm, 2008), in which a new vision about the classical influence-on metaphor has been proposed: Emotions do not simply influence a purely rational process, but they are virtually part of any decision-making process. Therefore, in our work we do not consider emotions as in the classical influence-on metaphor, i.e. as contextual factors (Deng et al., 2015), but we take into account their *information role* when comparing options. In fact, it is well recognized that affective states influence which strategy of information processing individuals are likely to adopt (Schwarz, 2000). For instance, individuals who are in a happy mood are more likely to adopt a heuristic processing strategy, with high reliance on past experience and relatively little attention to the details of the specific situation.

In this work, the term *affective state* refers to the experience of feeling the underlying emotional state defined by a set of emotions (e.g., joy, surprise, anger, etc.). Inspired by the ideas behind the NDM, we considered affective states as part of the decision-making process: They are associated with past experiences of the user, such as preference elicitation, hence they contribute to the user modeling and can be consulted when comparing past experiences. The idea is that the affective state felt when giving a preference could be a "signature" for that preference, which could be exploited in the user modeling and recommendation phases. Indeed a good evaluation of an item does not mean that the item is good no matter what. It means that the product fits the needs of this customer in that specific mood (Winoto and Tang, 2010). The immediate issue that arises is how to acquire the information that allow to define the affective state of a user. In the next section, we will analyze some solutions to this problem.

2.2. Affective aspects for personalization systems

Systems designed for providing personalized services are based on the collection of user preferences either through explicit feedback (e.g., questionnaires, preference ratings) or implicit feedback, by analyzing user behaviour when interacting with the system. Preferences are generally related only to the features of the items liked by the user, even though in the last decade contextual information has been integrated in the user model with the aim of increasing the accuracy of personalization. It is common to find systems that rely on a semantic description of the user environment (Adomavicius and Tuzhilin, 2011), the time of the day (Schilit, Adams, and Want, 1994; Li et al., 2014), the weather (Braunhofer et al., 2013) and many other aspects that are complementary to the decision-making process (Baldauf, Dustdar, and Rosenberg, 2007).

In (Poirson and Da Cunha, 2018) is validated the hypothesis that if emotions influence preference, then using emotional similarity improves the product recommendations. However, compared to our model this recommender system implements a collaborative-filtering model.

Some authors propose the use of personality aspects (Ferwerda et al., 2015), moods (Lee et al., 2011; Winoto and Tang, 2010) and, more generally, user affective states. Initially, the interest in affective aspects has been limited by the lack of methods for collecting them. Long questionnaires to complete was the only strategy that allowed to acquire accurate information about psychological aspects.

Nowadays, as a consequence of the availability of user information on social networks, new strategies of affect detection are proposed (Liang et al., 2018). They are generally based on text analysis and can be grouped in machine learning-based approaches, lexicon-based approaches, and hybrid methods (Medhat, Hassan, and Korashy, 2014; Zhang and Liu, 2017). Each technique can work on an entire document, or at a phrase, or chunk level. Moreover, strategies based on fixed lexicons (Shao et al., 2019) are not directly applicable due to the high variability of the words used, usually altered by slangs, compressions and emoticons (Polignano et al., 2017a). Despite the complexity of the problem, sentiment detection can be formalized as a classification task in which the text must be associated with the corresponding affective class. Like every classification task faced with supervised machine learning approaches, the availability of labeled datasets is a critical aspect to consider. The complexity of the task is usually lowered by adopting hybrid solutions (Shanahan, Qu, and Wiebe, 2006). Generally, the output of these approaches is a set of scores each of which indicates the association between the emotion and the input test.

For example, given the input text “I had a very relaxing weekend”, the output could be Joy 0.7, Anger 0.0, Sadness 0.0, Surprise 0.2, Fear 0.1, Disgust 0.0” (Polignano et al., 2017a). Naturally, the scores might change according to the algorithm adopted.

2.3. Affect-aware personalization systems

Personalization based on affective and cognitive aspects has been introduced by Picard, who defined the concept of *Affective Computing.* She described the idea of a computer which can perceive the user affective state and adjust its response properly:

> “Computers that will interact naturally and intelligently with humans need the ability to at least recognize and express affect” (Picard, 1997).

The affective computer is based on the idea that an intelligent system should be able to predict what the user might like in a specific situation according to his feelings. In fact, the user moods and emotions strongly influence the consumption of a specific item. Recommender Systems (RecSys) are tools for helping people to make better choices; they should be founded on a solid broad understanding of how people make choices and how the process of making choices can be supported (Jameson et al., 2015). From this perspective, RecSys have started to take into account not only preferences, but also user emotions (Tkalcic et al., 2012; Kaminskas and Ricci, 2016; Zheng, Mobasher, and Burke, 2016; de Gemmis, Lops, and Semeraro, 2016) and personality (Chen, Wu, and He, 2016) when computing suggestions, according to a holistic view of the user. In other approaches, ”likes” and ”co-likes” left by users about items are starting to be largely used to optimize the probability of suggesting relevant items for the user. An innovative approach that exploits this idea is proposed by Tran et al., 2018 that uses a novel Regularized Multi-Embedding (RME) approach for significantly improving Recall@5 in many recommendation scenarios. Another innovative approach is also proposed by Papadakis, Panagiotakis, and

Fragopoulou, 2017 that discusses SCoR Synthetic Coordinate based Recommendation system. SCoR is based on the Vivaldi algorithm that places items in Euclidean space in order to put closer items that the user could like.

Two important issues arise from the need to model those aspects of the decision process in RecSys:

1. When should RecSys collect affective information?
2. How can affective information be exploited by recommendation algorithms that take into account the ways in which affect actually influences the decision process?

In Recommender Systems literature, affective information is mainly associated with multimedia content (Soleymani, Pantic, and Pun, 2012; Soleymani and Pantic, 2012; Tkalcic, Burnik, and Kosir, 2010) and plays different roles related to the acquisition of user preferences:

1. As a source of affective metadata for item modeling and building a preference model (Tkalcic et al., 2013; Joho et al., 2011);
2. As an implicit relevance feedback for assessing user satisfaction (Arapakis, Konstas, and Jose, 2009; Arapakis, Athanasakos, and Jose, 2010; de Gemmis et al., 2015a; de Gemmis, Lops, and Semeraro, 2016).

In this work, we focus on the first issue. We could state that the acquisition of emotional feedback and the consequent association of items with affective metadata is similar in practice to the process by which the chooser gathers both experiences and feelings they induced. The exploitation of items the user experienced in the past and corresponding affective labels for user profiling and recommendation correspond, in the decision process, to the step in which the chooser recalls past experiences to have a direction for the current decision problem. Examples of this kind of situation are found in content-based RecSys literature (Tkalcic et al., 2013; Joho et al., 2011). The main issue addressed by these studies is the identification of a valid set of affective features that allows the definition of an effective user model for the canonical *relevant/non-relevant* item categorization. Similar studies are found in the domain of music recommendation, where mood-aware RecSys are commonly designed (Han et al., 2010; Park, Yoo, and Cho, 2006). These systems are based on strategies of music emotion recognition: items are annotated with emotional labels by exploiting physical characteristics of songs, such as frequency or beats per minutes (Kim et al., 2010; Yang et al., 2007; Kim and Andre, 2008). The main challenge from both a user modeling and decision-making perspective is how to represent the whole affective state of the user in terms of emotions, mood, and personality.

There is another line of Recommender Systems research that considers emotions and feelings as in the classical *influence-on* view. Affective information is not directly included in the preference model, but it is used as contextual variables by Context-aware RecSys (CARS) (Gonzalez et al., 2007; Zheng, Mobasher, and Burke, 2013; Zheng, Mobasher, and Burke, 2016; Shi, Larson, and Hanjalic, 2010). These systems try to adapt user preferences across different contexts (e.g. time, weather, company, etc.); the most common approach is to use emotional contextual variables to perform pre-filtering or post-filtering on items (Adomavicius and Tuzhilin, 2011). The pre-filtering approach reduces the space of items to those that are more appropriate for a given context. Conversely, the post-filtering approach performs the filtering after generating the recommendations. Gonzales was one of the first researchers who use cognitive aspects as domain descriptive features for CARS (González, López, and Rosa, 2004). In that work, a user model is defined to associate characteristics of an item to a finite set of emotional labels describing the polarity of the user mood. This labeling has been used both for identifying the item features that lead the user mood in a positive state, and as contextual dimensions for pre- or post-filtering.

Qian et al., 2019, Quian et al. use in a recommender system model the user rating data as explicit information, user social network data as implicit information and sentiment from user reviews as emotional information. The authors suppose that implicit feedback such as psychological aspects can increase the accuracy of recommendation by converting the recommended task into the probability of maximizing the user selection behavior. Our proposed approach similarly suggests to define a process able to maximize the probability that a specific item is relevant for the user in a specific affective state. Differently, from this proposed solution, we decided to use all the six Ekman emotions to represent both user and items and to base our algorithm on a content-based approach instead of a classic collaborative filtering one.

To conclude, among different approaches, we propose to map the affective state of the user into a vector space of emotions, by observing his behavior when eliciting a preference. A "static" emotional label is associated to an item, as in previously discussed research, but we included in the user profile also the affective state of the user when providing the preference on that item.

Accordingly, the main contributions of this paper are:

1. The definition of both an affective profile and a function for computing an affective coherence score that measures how much an item matches that profile;
2. The adaptation of a state-of-art recommendation model to include the affective coherence score into the recommendation process.

In the next section, we will provide a formal definition of affective profile and we suggest how it could be used to adapt the recommendation process.

3. An Emotion-aware Computational Model for Personalization Systems

We propose a general model that could be adopted by personalization systems that exploit affective information to better tailor services to users. In Figure 1, a user-centered computational model is proposed, which takes into account user tendencies, behaviors, and preferences. The left side of the picture describes the part of the model devoted to perform data collection: digital footprints left on social networks (posts, tweets, reviews, etc.) and user behavior, observed when interacting with the system or other sources, are analyzed to extract preferences. The collected information is exploited to build an *affective user profile* which includes preferences and a description of the affective state felt by the user at a certain time. A formal description of the affective user profile is provided later in the paper. The *Affective Coherence Model* exploits both the affective user profile and current user affective state at the time when personalization must be performed to adapt the interaction. For instance, if a list of items must be suggested to the active user, the model computes an affective coherence score for each item i in the catalog, that measures whether i is suitable for the current user affective state.

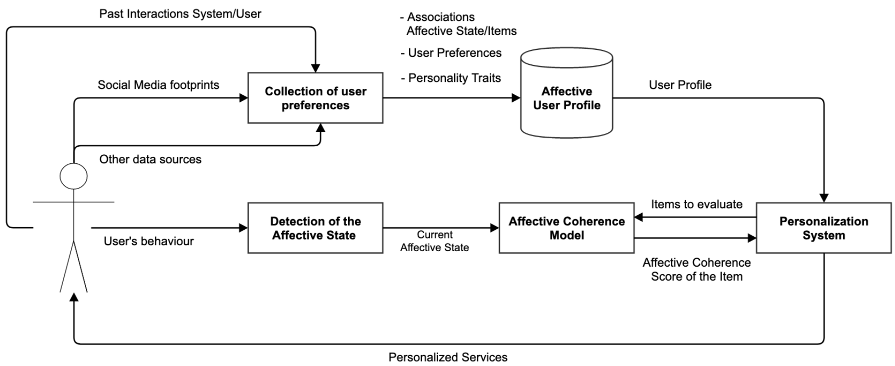

Figure 1: A General Affective-Aware computational model for personalized systems

Let I the set of items $\{i_1, i_2, ..., i_N\}$, u the *active user*, i.e. the user for whom the system will produce suggestions. The model is based on a list of preferences P_u collected during past interactions of u with the system. Each preference $p_{j,t} \in P_u$ is a rating r_j collected at time t on a specific item i_j:

$$p_{j,t} = \langle i_j, r_j \rangle \quad (1)$$

The affective state of u at time t is represented by an *affective vector* defined as:

$$\vec{s_t} = \langle e_1, e_2, e_3, e_4, e_5, e_6 \rangle \quad (2)$$

where each dimension value e_i is a weight which shows the presence of one of the six Ekman basic emotions (Joy, Anger, Sadness, Surprise, Fear, Disgust) (Ekman et al., 1987) in the user affective state.

The main advantage of this representation is that easy comparison between affective states can be computed, by adopting for instance simple cosine similarity.

In the same way as for preferences, also affective states can be collected when the user interacts with the system: $A_u = \{s_1, s_2, ...s_T\}$. The affective profile AP_u of user u stores both preferences P_u and affective states A_u, so that each collected preference $p_{j,t}$ can be associated, if possible, with the affective state s_t felt by u when that preference is provided:

$$AP_u = \langle P_u, A_u \rangle \tag{3}$$

In our model, we assume that also items are associated with emotions. For instance, an horror movie usually associated with fear or disgust. Obviously the emotions that the *consumption* of an item might induce in a specific user depend on several variables, even contextual: the same romantic song might cause happiness in one person, but sadness in another person, just because the latter has been left by the partner. When we refer to emotions associated with an item i_j, we assume that an affective vector $\vec{a_j}$ can be *permanently* associated with i_j, depending only on item features. Therefore, while a user could be associated with an affective vector each time he interacts with the system, the affective profile of an item is final. The strategy that builds affective vectors for items is domain dependent: lyrics and melody could be exploited for music recommendation, while genre or plot are useful for movies. In Section 5 we will show how to compute affective vectors for songs.

Given this representations for affective profiles of users and items, it is possible to define a function that computes, at time t', the affective similarity between item i_z, to be recommended to user u, and any past user preference $p_{j,t} \in P_u$:

$$affsim(i_z, p_{j,t}) = 1/2 * sim(\vec{a_z}, \vec{a_j}) + 1/2 * sim(\vec{s_{t'}}, \vec{s_t}) \tag{4}$$

In our view, the similarity equally depends on similarity between affective profiles of items and similarity between the current affective state of u and his affective state $\vec{s_t}$, felt when he provided the rating on item i_j.

The final output of the model is an *affective coherence score* which estimates how much item i_z is suitable for u, by taking into account both his current affective state at time t and his affective profile AP_u. Therefore, we define a new function, based on affective similarity defined in Equation 4, that allows comparison between i_z and *all* preferences stored in AP_u:

$$affcoh(i_z, AP_u) = \max_{p_{j,t} \in AP_u} affsim(i_z, p_{j,t}) \quad (5)$$

In other words, affective coherence score estimates the affective suitability of item i_j for a user as the maximum value of affective similarity over all preferences he expressed in the past.

In the proposed model, affective coherence is an additional information that can be exploited in many ways by recommendation algorithms. One possibility is to add the affective score as an additional descriptive feature for items, which can be used by the filtering algorithm in the same way as other features or with some boosting factor (Zheng, Mobasher, and Burke, 2016). Another possible approach is to adopt the affective score for pre- or post-filtering, as suggested by Adomavicius (Adomavicius and Tuzhilin, 2011).

We propose a simple way to compute a recommendation score that can be easily exploited by algorithms based on item similarity, such as content-based filtering (de Gemmis et al., 2015b) or item-based collaborative filtering (Sarwar et al., 2001). The idea is to compute an affective recommendation score that extends affective coherence defined by Eq. (5) by including also item similarity which does not depend on affective profiles of items:

$$aff\_rec\_score(i_z, AP_u) = \max_{p_{j,t} \in AP_u} (\alpha * affsim(i_z, p_{j,t}) + \beta * sim(i_z, i_j)) \quad (6)$$

where $sim(i_z, i_j)$ could be any similarity function, such as content similarity based on item features, or the number of users who co-rated the items.

Parameters α and β are used to weight differently affective and not-affective similarity. In the following section, we will show how an emotion-aware music RecSys can be implemented based on the proposed model.

4. EMRES: Emotion-aware Music REcommender System

EMRES computes a list of suggested songs for user u, by taking into account his affective profile. The part of the model that performs data collection is implemented by defining:

1. a strategy for collecting user affective states and preferences on songs;
2. a model to represent songs and to compute song similarity.

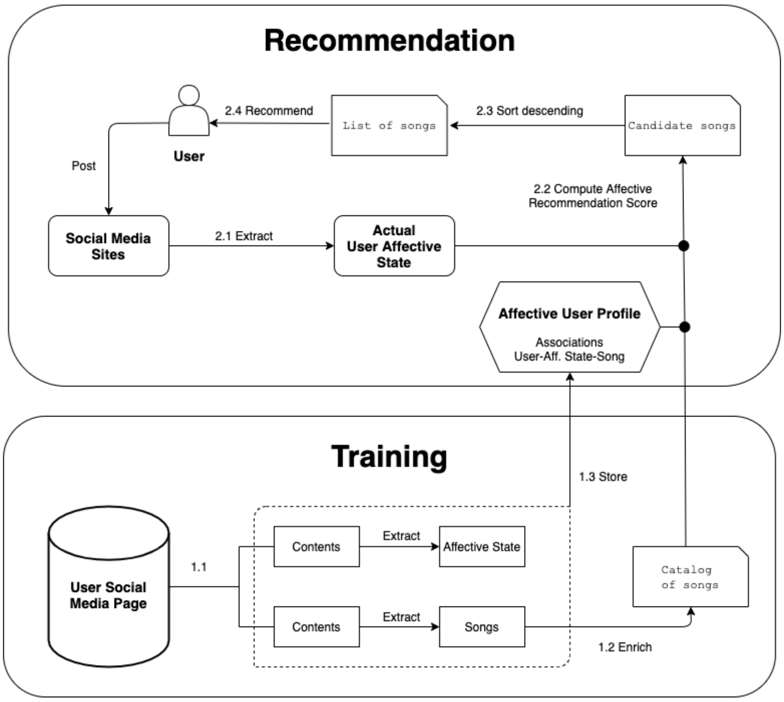

Figure 2: General diagram of the main steps of the EMRES model.

Figure 2 shows the general model of EMRES and the main procedural steps for obtaining recommendations. In particular, starting from point 1.1 of the training phase, it is possible to notice how we extract songs and textual contents from social media content. We focus on songs published through YouTube, and user messages wrote in a time laps around them. On this data, we perform sentiment analysis for extracting the user affective state and the emotions evoked by the song through its lyrics. Once the songs are derived from the social content, they are enriched with additional information, features of the audio file, TF-IDF representation of their lyrics, music genres. These elements are consequently saved in our recommendation catalog during step 1.2. At the same time, the User Affective Profile is also formalized in step 1.3. The songs that the user prefers to listen to in a specific emotional state are saved as user affective preferences. Once all background information is collected, it is possible to proceed next to the recommendation phase. Starting from step 2.1, the user's current affective status is extracted from her social contents. It, together with the User Affective Profile and the catalog, is used in phase 2.2. to calculate the Affective Recommendation score as per Eq. 6. The songs selected by this process are sorted in descending order by the score obtained and finally suggested to the user in phase 2.4. The steps just described will be better following detailed in this section.

4.1. Collecting affective states and preferences

For the detection of the affective state of the user u at a certain time t, we exploit textual messages posted by user u on Social Media Sites (SMSs) in the time window $W = [t-\kappa, t+\kappa]$, where κ is a parameter which defines the size of the time window. Let $M = \{m_1, m_2, ..., m_n\}$ be the set of n textual messages collected within W. Obviously the definition of κ is an issue: on one hand, choosing a small value for that parameter could allow to estimate the affective state of the user with good precision, but it might happen that M is empty. On the other hand, higher values for κ might result in a shallow estimate of the user affective state. Message $m_i \in M$ is processed by *Tone Analyzer*[1], the emotional labeling algorithm provided by IBM Bluemix, that directly provides the corresponding affective vector $\vec{m_i}$ as the distribution of Ekman emotions conveyed by m_i.

[1] https://www.ibm.com/watson/services/tone-analyzer/

In order to compute the affective state of u at time t, the n affective vectors associated with posts falling into the time window W are aggregated:

$$\vec{s_t} = \frac{\sum_{i=1}^{n} \gamma_i * \vec{m_i}}{\sum_{i=1}^{n} \gamma_i} \tag{7}$$

where an exponential decay factor γ_i, falling into the interval $[0, 1]$, is used to smooth the contribution of messages that are far from t:

$$\gamma_i = 1 - e^{\frac{-\Delta T}{\tau}} \tag{8}$$

Music preferences to be included in the affective user profile AP_u are collected from user posts on SMSs as well. Given a message m_i posted by u at time t, we search for the string pattern $*www.youtube.com/watch?v = *$ in order to detect whether m_i refers to a song i_j. For instance, the url $www.youtube.com/watch?v = EM4vblG6BVQ$ refers to a famous song by U2; we use the YouTube API to associate a category (music, sport, etc.) to the code $EM4vblG6BVQ$. In case it is recognized as a music link, a positive preference is included in AP_u for that song. In fact, in absence of explicit ratings, we assume that every song shared on SMSs is liked by the user. In our model, a traditional 5-stars rating scale is adopted. Continuing with our example, the preference $\langle EM4vblG6BVQ, 5\rangle$ is added to AP_u, together with the affective state s_t felt by u in the chosen time window, computed as in Eq. (7). We give the highest score for like, but other implicit feedback strategies could be adopted.

In the experiments, reported in Section 5, we will show how the proposed approach has been exploited to extract music preferences from two benchmark datasets.

4.2. Song Representation Model

A content-based approach is adopted to represent songs. Each song is associated with 3 sets of features:

1. low-level features, describing audio properties of the song;
2. music genres (e.g. rock, pop);

3. textual features extracted from lyrics.

Starting from the title of a song, 11 low-level audio features are collected from Spotify[2]:

- *Acousticness*: it is a confidence measure, ranging from 0.0 to 1.0, that tells whether the track is acoustic;
- *Danceability*: it describes how suitable a track is for dancing, based on a combination of musical elements, including tempo, rhythm stability, etc.
- *Energy*: it's a measure, ranging from 0.0 to 1.0, representing a perceptual measure of intensity and activity. Typically, energetic tracks are fast, loud, and noisy;
- *Instrumentalness*: Predicts whether a track contains no vocals. The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content;
- *Key*: The key of the track. Integers map to pitches using standard Pitch Class notation;
- *Liveness*: Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live;
- *Loudness*: The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Typical values range between -60 and 0 dB;
- *Mode*: it indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1, while minor is 0;

[2]http://www.spotify.com

- *Speechiness*: it detects the presence of spoken words in a track. Values below 0.33 most likely represent music and other non-speech-like tracks.

- *Tempo*: the overall estimated tempo of a track in beats per minute. In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration;

As for genres, we collected genre tags by using `Last.fm`[3] API. The problem with the tag list provided by `Last.fm` is that it is sometimes too long and contains specific tags which are not useful to understand music preferences (e.g. "Female Singer"). According to the approach proposed by Ferwerda (Ferwerda et al., 2015), tags have been filtered by using the *Allmusic*[4] list: *Alternative, Blues, Classical, Country, Electronic, Electronic, Folk, Jazz, Latin, New Age, Pop, Rock, Rap, R&B, Reggae, Stage & Screen, Vocal, World.* Each genre has been encoded as a binary feature of the song.

Textual features are collected by searching for the song title on Google[5] and Yahoo[6]. The result set has been bounded by the most widely used lyrics websites (Tab. A.11).

We took the top-10 results for each search engine, and we computed the *Jaro-Winkler* similarity (Jaro, 1989) between the search title, as provided by Youtube, and the titles in the result set, because it might happen that the title of a song has many variants on the web. We ranked the results according to *Jaro-Winkler* similarity score, and the top-ranked one has been chosen for collecting the lyrics, provided that the score exceeded 0.85. Otherwise, the song was discarded.

The lyrics is processed by basic NLP operations (stop-words removal, n-gram tokenization, and stemming) and $TF - IDF$ scores computed. Formally, each song i_j is associated with a vector which includes the following features:

[3]https://www.last.fm/
[4]http://www.allmusic.com
[5]http://www.google.com
[6]http://www.yahoo.com

$$\vec{i_j} = \langle TF\text{–}IDF\_lyric, low\_level\_features, genres\_tags \rangle \tag{9}$$

Cosine similarity between songs i_z and i_j can be computed between the two content vectors $\vec{i_z}$ and $\vec{i_j}$, in order to assess not-affective similarity by Eq. (6). In order to compute affective similarity, we need a way to associate affective vectors to songs. We processed the plain text of a song i_j by IBM Bluemix *Tone Analyzer* to obtain $\vec{s_j}$, in the same way as for user posts: $\vec{s_j}$ should describe the emotions that the author of the song wants to evoke in the listener. Given this representations for songs, both content vectors and affective vectors can be used to compute the affective recommendation score for song i_z, given the affective profile AP_u of the target user u, according to Eq. (6).

5. Experimental Evaluation

The main aim of the evaluation is to compare our emotion-aware RecSys EMRES with alternative approaches which do not take into account affective information, in order to show that the adoption of the proposed Affective Coherence Model improves the accuracy of suggestions. As alternative approaches we consider: (i) a content-based filtering algorithm that exploits the same non-affective features as our system; (ii) a recommendation method based on machine learning techniques for rating prediction. More specifically, the following research questions are formulated:

- **RQ1:** Does EMRES improve the accuracy of a classic content-based recommendation algorithm?
- **RQ2:** Does EMRES improve the accuracy of a classic content-based recommendation algorithm which exploits affective information for pre-filtering?
- **RQ3:** Does EMRES improve the accuracy of a recommendation algorithm based on logistic regression?

- **RQ4:** There is any configuration of parameters for computing the affective recommendation score that performs better than others?

RQ1, RQ2 and RQ3 refer to the baselines proposed for the comparison. In particular, while RQ1 refers to a "pure" content-based approach, RQ2 wants to investigate an alternative use of the affective coherence score, as suggested in (Zheng, Mobasher, and Burke, 2013). RQ3 investigates the comparison with an effective machine learning method for rating prediction, while RQ4 is formulated to understand how to balance the two similarities involved in the computation of the affective recommendation score.

5.1. Datasets and song catalogs

Experiments have been performed over two datasets: *myPersonality* (Kosinski et al., 2015) and *SNAP twitter7* (Yang and Leskovec, 2011). *myPersonality* contains information about 4 million Facebook users and 22 million posts over their timeline[7].

In our experiment, we considered only status update messages. Each message is stored as a triple *(userID, timestamp, published_text)*.

SNAP twitter7 is a very large dataset of tweets released by Stanford University. It contains 476 million messages, provided by 17 millions of users, collected through the Twitter API between June 1^{st}, 2009 and December 31^{st}, 2009. For each tweet, *'id'* of the user, the timestamp and the textual message are available. Also for this dataset, we considered only status update messages, which are stored as triples *(userID, timestamp, published_tweet)*.

In order to have reliable data for learning preferences to be included in the user affective profile, described by Eq. (3) in Section 3, we considered only users who:

1. published at least 4 music links;

[7] http://web.archive.org/web/20180428085709/mypersonality.org/wiki/doku.php?id=list˙of˙variables˙available

2. posted at least another message (not a music link) around the timestamp of a published music link.

Other messages posted at a time close when the user shared a music link are needed to build his affective state, that will be associated with the music preference in the affective profile. Table 1 shows the information stored in both datasets after this filtering phase. On average, about 5 songs per user are available. It is important to point out that this filtering step significantly reduced the volume of data, which has not made it possible the comparison with a collaborative algorithm.

Table 1: Description of the datasets.

	myPersonality	**SNAP**
# Tot. Users	4M	17M
# Tot. Posts	22M	476M
# Users who posted music links	109	1,124
# Music links	509	5,837
Avg # Music links per User	4.67	5.19

Given a triple *(userID, timestamp, published_text)*, if *published_text* is recognized as a music link by the strategy described in Section 4.1, content features of the corresponding song are collected in order to build both a content vector and an affective vector for that song. In Section 4.2, we already provided the details of the song representation model: lyrics have been turned into $TF - IDF$ vectors ($7,693$ features for *myPersonality*, $35,677$ features for *SNAP twitter7*); 11 low-level features have been extracted from Spotify, and genre tags have been collected by using `Last.fm` API. For each song, a content vector is built, as shown in Eq. (9). Furthermore, an affective vector for the song is built by *IBM Tone Analyzer*. At the end of this process, we obtained two song catalogs that contain both content vectors and affective vectors for 509 *myPersonality* songs, and for $5,837$ *SNAP twitter7* songs.

5.2. Metrics

Relevance in the context of recommendation is a user-specific notion which can be equated to the interest of users for items (Castells, Hurley, and Vargas, 2015) and can be modeled as a binary concept: either an item is liked by a user or not. According to this idea, an item i could be defined as relevant to user u if the rating given by u on i is greater than the average value of all ratings provided by u.

When considering relevance depending also on the affective state of the user, the classification of an item as relevant or not relevant is not defined in such a sharp way as when it is based on the user average rating. Our approach is not based on rating estimation, but it computes an affective recommendation score that is useful to rank items according to their suitability with both the affective profile of the user and his current affective state. Therefore, we are interested in assessing the capability of our system to place this kind of items in the top-ranked positions of the recommendation list. For this reason, we preferred *HitRate* ($HR@n$) to the classic classification metrics. For a single user, the metric is computed as follows:

- one target item i is randomly picked up from the test set;
- the system computes n recommendations;
- if item i appears in the n recommendations, it is a hit.

The whole $HR@n$ of the system is the count of hits, divided by the test user count. In our experiments, we choose $n = 1, 5, 10, 15, 20, 100$. The motivation behind the choice of these values is to simulate different recommendation scenarios (single song or playlists of different sizes).

5.3. Evaluation Protocol for EMRES

Experiments were carried out using a per user evaluation, scheduled as follows:

1. For each user u in the dataset, triples are randomly distributed into the training set T_r (70%) and the test set T_s (30%).

2. The affective profile AP_u is built. For each triple *(u, t, l_j)* $\in T_r$:

   (a) preference $p_{j,t}$ for the music link l_j is included in P_u:

$$p_{j,t} = \langle l_j, r_j = 5 \rangle \tag{10}$$

   (b) affective vector $\vec{s_t}$ for u is computed by Eq. (7), so that triple *(u, t, l_j)* is associated with the affective state felt by the user when his preference for l_j was elicited. The affective vector is built by averaging the affective vectors associated by *IBM Bluemix Tone Analyzer* to messages falling into a 24 hours time window $[t-12, t+12]$ for *myPersonality* and a 12 hours window $[t-6, t+6]$ for *SNAP twitter7*. The different time windows were due to the different sizes of the datasets. Affective vector $\vec{s_t}$ is included in A_u.

3. For each *target* triple *(u, t', l_z)* $\in T_s$:

   (a) affective vector $\vec{s_{t'}}$ is built, in the same way as for the training set;

   (b) vector $\vec{s_{t'}}$ is used to compute a recommendation list L suitable for that affective state: all items i_k in the song catalog (not belonging to T_r) are ranked according to the affective recommendation score $aff\_rec\_score(i_k, AP_u)$, computed by Eq. (6). Three different settings are chosen for parameters α and β: (i) $\alpha = 0.8$, $\beta = 0.2$; (ii) $\alpha = 0.5$, $\beta = 0.5$; (iii) $\alpha = 0.2$ and $\beta = 0.8$.

   (c) $HR@n$ is computed on L for the target item l_z, as described in the previous section.

4. $HR@n$ is averaged on $|T_s|$ and then on the number of users.

5.4. Compared algorithms

We compared EMRES to the following algorithms:

- **Classic content-based recommendation algorithm (CB).** Each song is represented by the content vector $\vec{i_j}$, as defined by Eq. (9). The evaluation protocol is the same as for EMRES but, given the target user u, all items i_k in the song catalog (not belonging to T_r) are ranked according to the affective recommendation score $aff\_rec\_score(i_k, AP_u)$, computed as defined by Eq. (6) in Section 3, where $\alpha = 0$ and $\beta = 1$. Therefore, the affective component of the user profile does not contribute to the similarity computation, which is based only on cosine similarity between content features of songs.

- **Classic content-based recommendation algorithm with Pre-Filtering (CBPF).** This baseline has been adopted to investigate RQ2. This baseline is inspired by the user-splitting method proposed in (Zheng, Mobasher, and Burke, 2016) and it is adopted in order to compare EMRES with a basic method that exploits the affective state as contextual filtering factor. The evaluation protocol has been adapted to filter out the items which are not suitable for the current affective state of the user. For each *target* triple *(u, t', l_z)* $\in T_s$, the predominant affective state of u is defined as the emotion e associated with the maximum value in the emotional vector $\vec{s_{t'}}$ (e.g. sadness). Pre-filtering is applied on the catalog in order to select only the songs whose predominant emotion coincides with e, by exploiting the affective vectors associated with songs. Therefore, in our example, the output of the pre-filtering step is the subset of "sad" songs within the catalog. Then, *CB* is applied to rank songs as in the previous baseline.

- **Logistic Regression (L1-LR).** This kind on machine learning algorithm is widely used for many classification problems, particularly ones with many features. We adopted *L1-regularized logistic regression*[8] with standard parameters (Lee et al., 2006). We treated the problem as a binary classification task: *Liked-Songs* vs. *Disliked-Songs*. Given the target user

[8] https://www.csie.ntu.edu.tw/~cjlin/liblinear/

u, we included in *Liked-Songs* all the songs posted by u. Then, we added 20 songs selected by a k-NN algorithm based only on content features, with the aim of increasing the low number of training instances. In order to have balanced classes, $|$*Liked-Songs*$|$ songs have been included in *Disliked-Songs*, randomly selected from the catalog. The algorithm is trained with 70% of the songs available in the two sets. The class probability estimation for *Liked-Songs*, computed by the classification model on test items, is adopted as a ranking score to build the recommendation list.

- **EARS-Qian (EARS).** The approach want to be a baseline for comparing our proposed approach with it developed by Qian et al., 2019. In particular we are forced to partially modify it in order to make it compliant with our dataset. Considering the information fusion algorithm proposed y the authors the recommendation algorithm should take into account the social media information about the user, the explicit feedback and the implicit feedback. In our approach to the problem we are working only with implicit information such as the action of sharing a song on social media, the emotions evoked by the song and the user affective state at the moment of interaction with the system. Moreover in absence of explicit feedback we can only apply a recommendation strategy based on contents. For this reason we are mapping the Quian's approach as follow:

  - User's affcetive state s_t and the item evoked emotion a_j is considered as a polarity score value between 0 and 1 using measured using SentiWordNet 3.0. More details about the process are described into the article of Qian et al., 2019.

  - For each item i_j in the User Profile, we are selecting the most similar songs in catalog using the score obtained by applying a cosine similarity among their item vectors, as described in Eq. 9.

  - The similarity scores among a catalog item and a previous user preference is influenced by an exponential decay time factor as described in Eq. 8. More the user preference is old less is its importance in selecting the recommendation candidate songs.

– The score obtained by the previous step is multiplied by the affective score of s_t, a_j.

– The items are ordered descending by their scores; duplicates are removed, leaving only the first occurrence of it; a set of k elements are returned as EARS recommendations.

Since we have different kinds of content features (low-level audio features, music genres, textual features), we performed several runs of the experiments, each one using a different kind of features.

5.5. Analysis of results and Discussions

The results on *myPersonality* are reported in Tables 2 - 5. Each table reports the $HR@n$ values obtained by using a specific combination of features. For each size of recommendation list, the highest *HitRate* is marked in bold. Since we have 3 different configurations for EMRES, we marked with $'*'$ the best result, only whether the difference with other baselines is statistically significant, by using Wilcoxon Signed-Rank Test two-tails ($p < 0.05$). As a consequence of the multiple statistical comparisons we applied the Bonferroni correction.

The main outcome from Table 2 is that there is at least one configuration of EMRES that performs better than the baselines for each size of recommendation list. **EMRES-3**, which gives more weight to affective similarity, is the best system for $n \geq 10$, while affective similarity seems to have a more limited impact on shorter recommendation lists. The EARS model performs better than the baselines but worse than our proposed approach. This observation can suggest to us that the emotional component is positively influencing a classic content-based recommendation approach. Still, the absence of explicit feedback about the items, an essential element in the original Qian et al., 2019 model, affects the recommendation approach and the consequent results obtained negatively. Another relevant observation is that using affective information for pre-filtering

Table 2: *HitRate* on *myPersonality*. Item descriptions include all the available content features of the songs: Low level features, TF-IDF, genres.

Algorithm	**HR@1**	**HR@5**	**HR@10**	**HR@15**	**HR@20**	**HR@100**
CB	**0.018**	0.036	0.048	0.067	0.073	0.218
CBPF	0.006	0.024	0.042	0.048	0.048	0.218
L1-LR	0.012	0.012	0.024	0.030	0.036	0.134
EARS	0.018	0.056	0.087	0.107	0.121	0.294
EMRES-1 $\alpha = 0.2$, $\beta = 0.8$	0.012	**0.067**	0.091	**0.115**	0.134	0.370
EMRES-2 $\alpha = 0.5$, $\beta = 0.5$	**0.018**	0.060	0.096	**0.115**	**0.139**	0.376
EMRES-3 $\alpha = 0.8$, $\beta = 0.2$	0.012	0.048	**0.103**	**0.115**	**0.139**	**0.388***

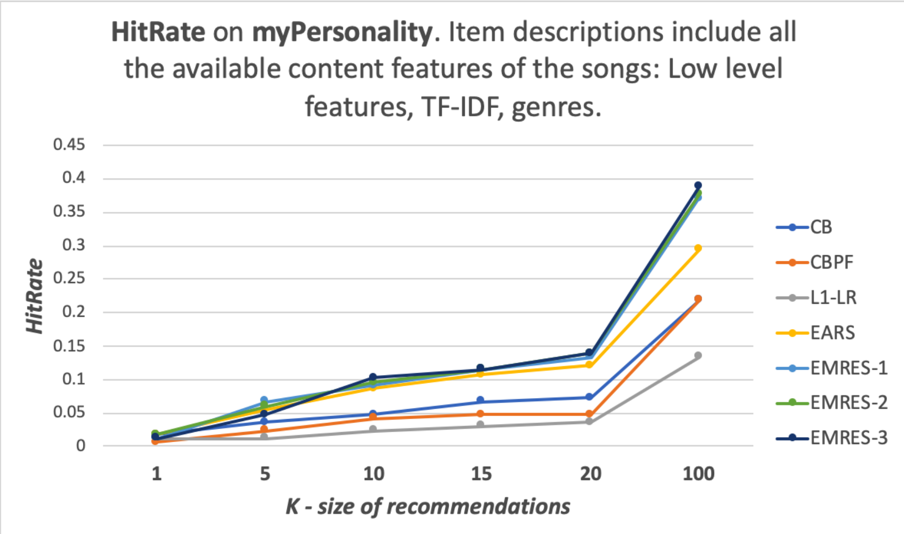

Figure 3: Graphical view of results showed in Table 2

worsens the performance of the content-based algorithm, which anyway outperformed linear regression, whose poor results are due to the limited size of the training set.

By using only low level features (Table 3), we observed that **EMRES** outperformed the compared algorithms as well. As for the different configuration

Table 3: *HitRate* on *myPersonality*. Item descriptions include only low level features.

Algorithm	HR@1	HR@5	HR@10	HR@15	HR@20	HR@100
CB	**0.012**	0.024	0.024	0.055	0.055	0.176
CBPF	0.006	0.024	0.042	0.048	0.048	0.218
L1-LR	**0.012**	0.018	0.024	0.030	0.042	0.158
EARS	**0.012**	0.038	0.072	0.096	0.0115	0.326
EMRES-1 $\alpha = 0.2$, $\beta = 0.8$	**0.012**	0.042	**0.090***	**0.121**	**0.145***	**0.406***
EMRES-2 $\alpha = 0.5$, $\beta = 0.5$	**0.012**	0.042	**0.090**	0.115	0.139	0.381
EMRES-3 $\alpha = 0.8$, $\beta = 0.2$	**0.012**	**0.060**	**0.090**	0.115	**0.145**	0.381

of our method, it seems that **EMRES-1** obtained better results, because only its $HR@5$ value is below **EMRES-3**.

Figure 4: Graphical view of results showed in Table 3

By using only TF-IDF (Table 4), in general **EMRES-3** showed better results than compared algorithms. By looking also at Table 5, it is quite surprising that **CB**, **EARS** and **EMRES** improved their performance by using only

Table 4: *HitRate* on *myPersonality*. Item descriptions include only TF-IDF.

Algorithm	**HR@1**	**HR@5**	**HR@10**	**HR@15**	**HR@20**	**HR@100**
CB	0.030	0.042	0.079	0.121	0.145	0.334
CBPF	0.006	0.024	0.042	0.048	0.048	0.218
L1-LR	0.012	0.012	0.024	0.036	0.042	0.145
EARS	**0.036**	0.042	0.090	0.145	0.153	0.388
EMRES-1 $\alpha = 0.2$, $\beta = 0.8$	**0.036**	0.055	0.109	0.115	0.139	0.370
EMRES-2 $\alpha = 0.5$, $\beta = 0.5$	**0.036**	0.055	**0.115**	0.127	0.139	0.388
EMRES-3 $\alpha = 0.8$, $\beta = 0.2$	**0.036**	**0.073**	**0.115**	**0.164**	**0.170**	**0.394**

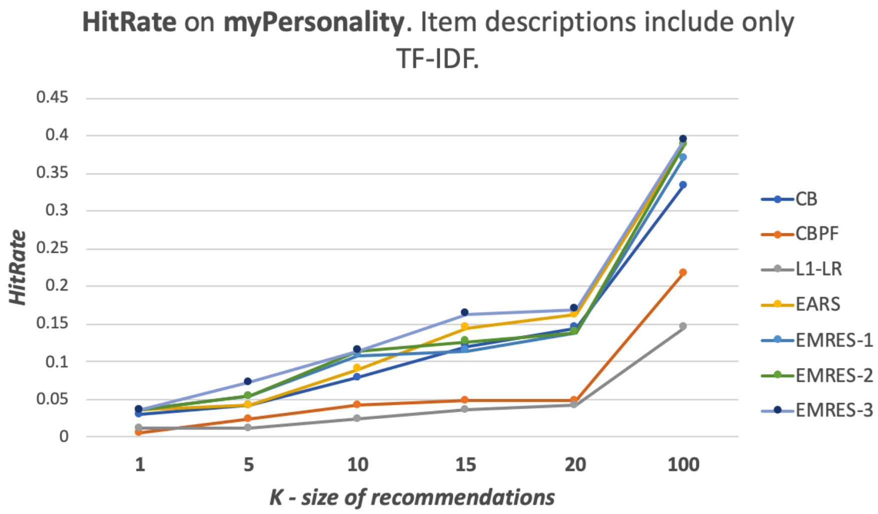

Figure 5: Graphical view of results showed in Table 4

textual features extracted from lyrics (thus neglecting both genres and audio features). This outcome could suggest that this kind of features is more useful than others to elicit music preferences whether a few training data are available. Furthermore, with music genres, **EMRES-1** seems the best configuration for our system. In summary, the results allow to answer positively only to **RQ1**,

RQ2, and **RQ3**, while there are conflicting results for **RQ4**, because it is not possible to identify a specific configuration of **EMRES** which performs better than others.

Table 5: *HitRate* on *myPersonality*. Item descriptions include only music genres.

Algorithm	**HR@1**	**HR@5**	**HR@10**	**HR@15**	**HR@20**	**HR@100**
CB	0.012	0.048	0.067	0.085	0.097	0.273
CBPF	0.006	0.030	0.030	0.042	0.042	0.218
L1-LR	0.012	0.018	0.024	0.036	0.036	0.134
EARS	0.018	0.054	0.070	0.090	0.097	0.285
EMRES-1 $\alpha = 0.2$, $\beta = 0.8$	**0.024**	**0.072**	**0.090**	**0.115**	**0.115**	0.309
EMRES-2 $\alpha = 0.5$, $\beta = 0.5$	**0.024**	0.067	**0.090**	0.109	**0.115**	0.327
EMRES-3 $\alpha = 0.8$, $\beta = 0.2$	0.018	0.054	0.073	0.090	0.103	**0.358***

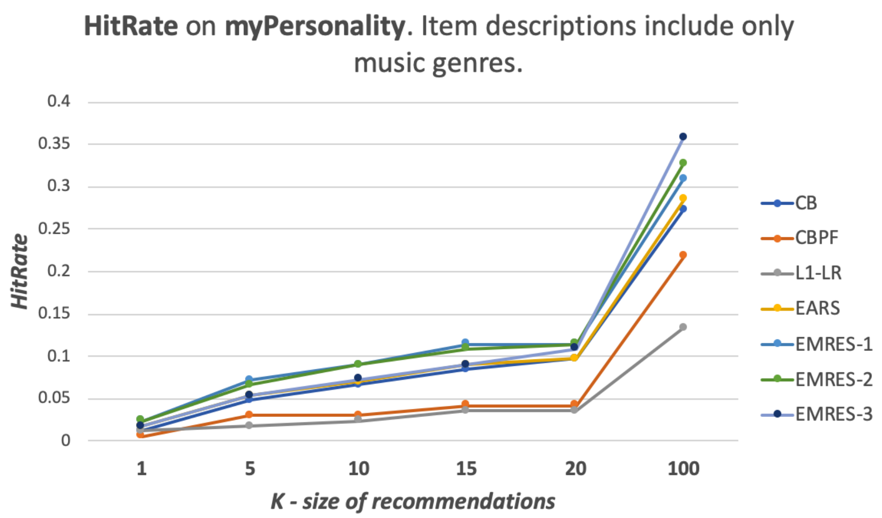

Figure 6: Graphical view of results showed in Table 5

The same experimental runs have been performed on *SNAP twitter7*, in which both the number of users and items is larger than in *myPersonality*. The results, reported in Tables 6 - 9, confirm the trend observed on *myPersonality*: **EMRES** outperformed the compared algorithms, and the differences are statistically significant, by using Wilcoxon Signed-Rank Test two-tails test ($p < 0.05$) and Bonferroni correction. For instance, in Table 6 and Table 7, there is a clear primacy of **EMRES**, but there is no configuration that prevails over others. Results in Table 8 are similar to those reported in Table 4: by using only TF-IDF, **EMRES-3** showed better results than **EMRES-1**, while **EMRES-2** prevails only for HR@10 and HR@15. Furthermore, also on this dataset, we observed higher performance of **EMRES-1** when using only music genres 9.

Table 6: *HitRate* on *SNAP twitter7*. Item descriptions include all the available content features of the songs: Low level features, TF-IDF, genres.

Algorithm	**HR@1**	**HR@5**	**HR@10**	**HR@15**	**HR@20**	**HR@100**
CB	0.007	0.009	0.011	0.012	0.013	0.021
CBPF	0.001	0.002	0.003	0.004	0.006	0.009
L1-LR	0.002	0.003	0.004	0.005	0.006	0.020
EARS	0.009	0.018	0.033	0.057	0.074	0.116
EMRES-1 $\alpha = 0.2$, $\beta = 0.8$	**0.019***	**0.062***	**0.084***	0.098	0.116	0.228
EMRES-2 $\alpha = 0.5$, $\beta = 0.5$	0.018	0.050	0.074	0.102	**0.120***	**0.231***
EMRES-3 $\alpha = 0.8$, $\beta = 0.2$	0.018	0.054	0.082	**0.104***	0.115	0.230

Moreover, we performed an ablation test on the low level features configuration to observe the variations in terms of system performance. In particular, we used the configuration of EMRES-3 that uses $\alpha = 0.8$, $\beta = 0.2$ on the dataset SNAP twitter7 by using only low level features for the item descriptions (Table 7). It has been observed that the features ”key” and ”loudness” produce noise in the similarity score leading to a worsening of the hit rate. By removing them, EMRES gets an *improvement* in terms of HitRate@100 of approximately 0.03 (on average) compared to the basic configuration. Similar results are observed for the other list sizes. On the contrary, by removing the feature “tempo” it has

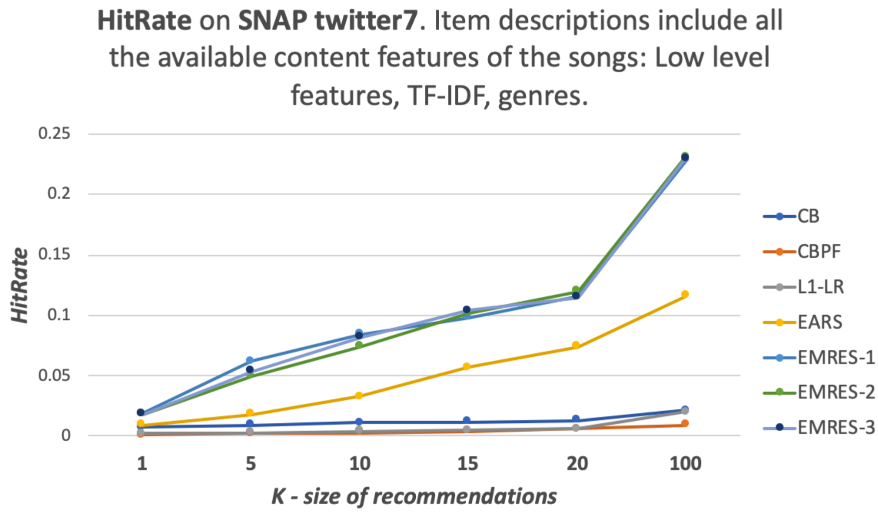

Figure 7: Graphical view of results showed in Table 6

Table 7: *HitRate* on *SNAP twitter7*. Item descriptions include only low level features.

Algorithm	**HR@1**	**HR@5**	**HR@10**	**HR@15**	**HR@20**	**HR@100**
CB	0.007	0.011	0.013	0.014	0.014	0.032
CBPF	0.000	0.001	0.003	0.003	0.005	0.020
L1-LR	0.000	0.001	0.003	0.006	0.009	0.014
EARS	0.009	0.021	0.033	0.062	0.085	0.132
EMRES-1 $\alpha = 0.2$, $\beta = 0.8$	**0.021***	0.057	0.079	0.098	0.117	0.278
EMRES-2 $\alpha = 0.5$, $\beta = 0.5$	0.018	**0.058***	0.082	0.105	**0.122***	**0.292**
EMRES-3 $\alpha = 0.8$, $\beta = 0.2$	0.017	0.055	**0.084***	**0.107***	0.121	**0.292***

been observed a *loss* of approximately 0.12 for the same metric. These differences are statistically significant according to the Wilcoxon Signed-Rank Test two-tails at $p < 0.05$.

In summary, the results confirmed that the proposed method performs better than both a content-based approach, even with pre-filtering based on the user's

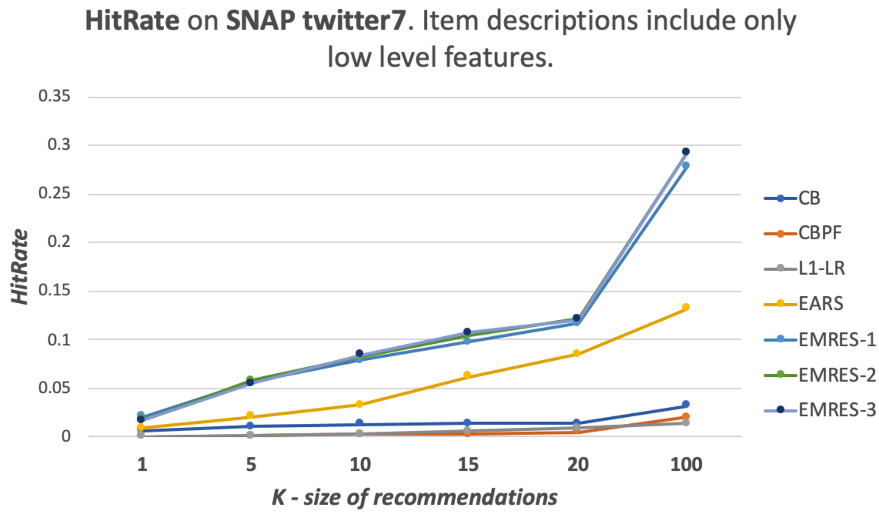

Figure 8: Graphical view of results showed in Table 7

Table 8: *HitRate* on *SNAP twitter7*. Item descriptions include only TF-IDF.

Algorithm	**HR@1**	**HR@5**	**HR@10**	**HR@15**	**HR@20**	**HR@100**
CB	0.010	0.016	0.021	0.024	0.025	0.045
CBPF	0.000	0.002	0.003	0.003	0.005	0.020
L1-LR	0.002	0.006	0.006	0.014	0.020	0.032
EARS	0.013	0.036	0.048	0.073	0.089	0.144
EMRES-1 $\alpha = 0.2$, $\beta = 0.8$	0.022	0.059	0.080	0.101	0.121	0.293
EMRES-2 $\alpha = 0.5$, $\beta = 0.5$	0.023	0.075	**0.090***	**0.116***	0.128	0.298
EMRES-3 $\alpha = 0.8$, $\beta = 0.2$	**0.024***	**0.077***	0.082	0.115	**0.132***	**0.306***

prevailing emotion within his affective state, and a classification approach based on logistic regression. Therefore, the main outcome is that emotional aspects are useful to adapt recommendations to the user affective state, but there is no clear indication on how to weight emotional factors when computing similarity between user preferences and item features. The main limitations of this model

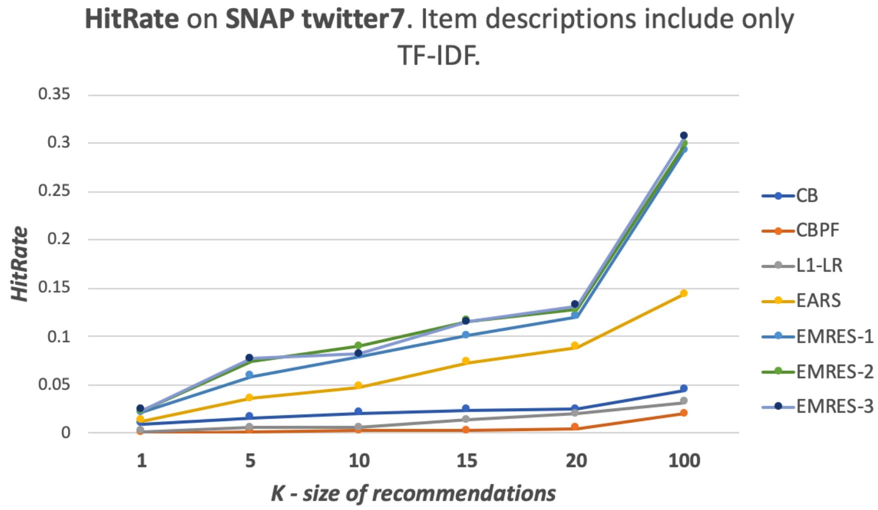

Figure 9: Graphical view of results showed in Table 8

Table 9: *HitRate* on *SNAP twitter7*. Item descriptions include only music genres.

Algorithm	**HR@1**	**HR@5**	**HR@10**	**HR@15**	**HR@20**	**HR@100**
CB	0.006	0.014	0.018	0.021	0.023	0.048
CBPF	0.000	0.002	0.003	0.004	0.005	0.022
L1-LR	0.000	0.004	0.007	0.011	0.018	0.032
EARS	0.009	0.016	0.022	0.034	0.043	0.087
EMRES-1 $\alpha = 0.2$, $\beta = 0.8$	0.013	**0.038***	**0.054***	**0.069***	**0.084***	**0.230***
EMRES-2 $\alpha = 0.5$, $\beta = 0.5$	0.013	0.037	0.051	0.065	0.081	0.224
EMRES-3 $\alpha = 0.8$, $\beta = 0.2$	**0.014***	0.037	0.053	0.067	0.082	0.224

is that it is only exploitable when textual content associated to the item is available. Furthermore, it requires that the user must have an active social profile in order to extract her prevalent emotions. Cold start and rating sparsity are common problems in many recommendation models. The cold-start problem is present in our model, as we stated before. The decision to use data from Social

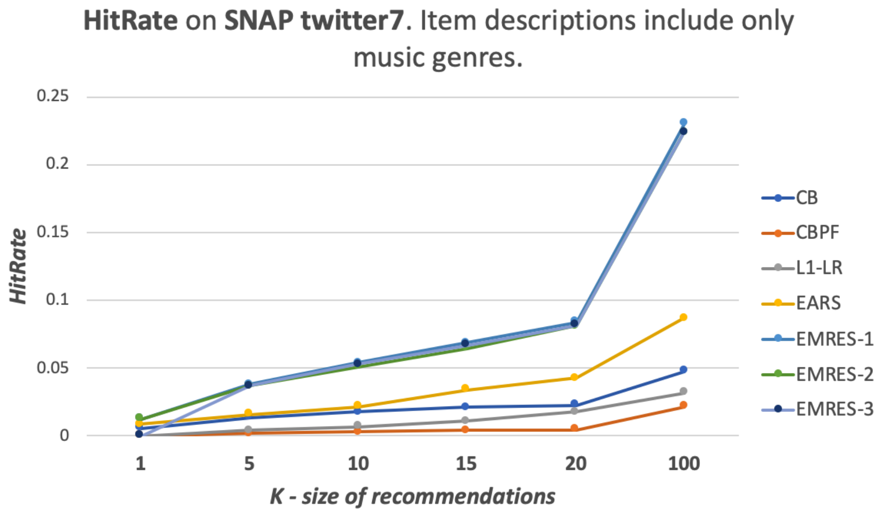

Figure 10: Graphical view of results showed in Table 9

Media as starting point of our model has exactly the aim of alleviating this problem, allowing the system to start from a preliminary set of user information. In case no social content is available for the user, a partial solution to this problem is to have an initial set of items to vote for each of the affective states in which the user might be. As far as the problem of rating sparsity is concerned, it is alleviated by using a content-based and not collaborative recommendation approach as the basis of EMRES. This indicates that the user's ratings alone, although few, are sufficient to obtain an initial recommendation set. Another limitation, related to the experimental evaluation, is that we haven't been able to identify the best parameter configurations for our model (RQ4). Hence, the system requires a preliminary tuning-parameter step before its use.

6. Conclusion

Traditional recommendation approaches have started to include novel aspects that describe in a deeper way the context in which a decision is taken, in order to increase both accuracy and the acceptance of proposed suggestions.

In this work, we investigated how to model the affective dimension of the user in a recommendation process, based on the idea that preferences of users might vary according to what they feel.

The main contribution of this work is a model that computes the affective coherence of an item for a user being in a specific affective state. Affective information is gathered from social media footprints, i.e. messages left on social networks, and used to model the user affective state as a vector in the space of Ekman's emotions. We defined both an affective profile and a function for computing an affective coherence score that measures how much an item matches that profile. Furthermore, we implemented a music recommender that adopts a content-based filtering method, extended with our affective coherence model. One problem with emotion-aware recommendation approaches is the lack of datasets for an in-vitro evaluation. In particular, it was challenging to find datasets containing all the information required by the proposed model to build affective user profiles. In order to have reliable data for learning user preferences to be associated with his affective state, in our evaluation protocol we exploited only a subset of users extracted from the two datasets *myPersonality* and *SNAP twitter7*.

The results of the investigation showed the effectiveness of the proposed approach, as a statistical significant increase of *Hit Rate* compared to: (a) a content-based recommendation algorithm that does not exploit affective information; (b) the same content-based method, which exploits affective information for pre-filtering of items; (c) a logistic regression classification method that learns user preferences from both content and affective features.

In particular, an interesting outcome is that pre-filtering based on the user's prevailing emotion, within his current affective state, worsens the performance of the content-based method. Maybe the reason for this result is that recommending items which reinforce the current emotional state of the user is a strategy that does not work for all users. For instance, while some users who feel sad prefer to listen to music that confirm their negative mood, other want to change it. The idea behind the proposed approach is that the way preferences depend on user emotional state varies from user to user.

As a future work, we would like to investigate whether user's inclination to empathy is another dimension that should be taken into account when learning user preferences by an emotion-aware approach. The motivation for this further study is that the prediction of a user's level of empathy could be considered as an indicator of how much emotions affect his preferences and decision-making processes. A preliminary study is described in (Polignano et al., 2017b). Moreover, we would like to improve the performances of the actual EMRES model by studying more strategies for accurately collecting the user emotions and opinions. In particular, about music tracks, we would like to investigate the possibility of monitoring the effective listening of a song, in order to exclude those elements just listened for a second and quickly skipped. Furthermore, we are planning to perform an *in-vivo* evaluation to overcome the difficulties due to the lack of benchmark datasets.

Acnowledgments

The research has been partially funded by the Apulia Region, Italy, POR Puglia FESR-FSE 2014−2020 Innonetwork, project DECiSION: DATA-DRIVEN CUSTOMER SERVICE INNOVATION.

References

Adomavicius, Gediminas and Alexander Tuzhilin (2011). "Context-aware recommender systems". In: *Recommender Systems Handbook*. Springer, pp. 217–253.

Alm, Cecilia Ovesdotter, Dan Roth, and Richard Sproat (2005). "Emotions from text: machine learning for text-based emotion prediction". In: *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 579–586.

Arapakis, Ioannis, Konstantinos Athanasakos, and Joemon M. Jose (2010). "A Comparison of General vs Personalised Affective Models for the Prediction of Topical Relevance". In: *Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2010, Geneva, Switzerland, July 19-23, 2010*. Ed. by Fabio Crestani, Stéphane Marchand-Maillet, Hsin-Hsi Chen, Efthimis N. Efthimiadis, and Jacques Savoy. ACM, pp. 371–378. DOI: 10.1145/1835449.1835512.

Arapakis, Ioannis, Ioannis Konstas, and Joemon M. Jose (2009). "Using Facial Expressions and Peripheral Physiological Signals as Implicit Indicators of Topical Relevance". In: *Proceedings of the 17th International Conference on Multimedia 2009, Vancouver, British Columbia, Canada, October 19-24, 2009*. Ed. by Wen Gao, Yong Rui, Alan Hanjalic, Changsheng Xu, Eckehard G. Steinbach, Abdulmotaleb El-Saddik, and Michelle X. Zhou. ACM, pp. 461–470. DOI: 10.1145/1631272.1631336.

Asur, Sitaram and Bernardo A Huberman (2010). "Predicting the Future with Social Media". In: *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*. Vol. 1. IEEE, pp. 492–499. DOI: 10.1109/WI-IAT.2010.63.

Back, Mitja D, Juliane M Stopfer, Simine Vazire, Sam Gaddis, Stefan C Schmukle, Boris Egloff, and Samuel D Gosling (2010). "Facebook profiles reflect actual personality, not self-idealization". In: *Psychological Science* 21.3, pp. 372–374. DOI: 10.1177/0956797609360756.

Baldauf, Matthias, Schahram Dustdar, and Florian Rosenberg (2007). "A survey on context-aware systems". In: *IJAHUC* 2.4, pp. 263–277. DOI: 10.1504/IJAHUC.2007.014070.

Bechara, Antoine (2003). "Risky business: emotion, decision-making, and addiction". In: *Journal of Gambling Studies* 19.1, pp. 23–51. DOI: 10.1023/A:1021223113233.

Bechara, Antoine, Hanna Damasio, and Antonio R Damasio (2000). "Emotion, decision making and the orbitofrontal cortex". In: *Cerebral Cortex* 10.3, pp. 295–307.

Braunhofer, Matthias, Mehdi Elahi, Francesco Ricci, and Thomas Schievenin (2013). “Context-aware points of interest suggestion with dynamic weather data management”. In: *Information and Communication Technologies in Tourism*. Springer, pp. 87–100. DOI: 10.1007/978-3-319-03973-2_7.

Burke, Robin D. (2002). “Hybrid Recommender Systems: Survey and Experiments”. In: *User Modeling User-Adapted Interaction* 12.4, pp. 331–370. DOI: 10.1023/A:1021240730564.

Castells, Pablo, Neil J. Hurley, and Saul Vargas (2015). “Novelty and Diversity in Recommender Systems”. In: *Recommender Systems Handbook*. Ed. by Francesco Ricci, Lior Rokach, and Bracha Shapira. Springer, pp. 881–918. DOI: 10.1007/978-1-4899-7637-6_26.

Chen, Li, Wen Wu, and Liang He (2016). “Personality and Recommendation Diversity”. In: *Emotions and Personality in Personalized Services: Models, Evaluation and Applications*. Ed. by Marko Tkalčič, Berardina De Carolis, Marco de Gemmis, Ante Odić, and Andrej Košir. Cham: Springer International Publishing, pp. 357–376. ISBN: 978-3-319-31413-6. DOI: 10.1007/978-3-319-31413-6_11.

Clore, Gerald L, Norbert Schwarz, and Michael Conway (1994). “Affective causes and consequences of social information processing”. In: *Handbook of Social Cognition* 1, pp. 323–417.

Correa, Teresa, Amber Willard Hinsley, and Homero Gil De Zuniga (2010). “Who interacts on the Web? The intersection of users’ personality and social media use”. In: *Computers in Human Behavior* 26.2, pp. 247–253. DOI: 10.1016/j.chb.2009.09.003.

Cunningham, Michael R (1988). “What do you do when you’re happy or blue? Mood, expectancies, and behavioral interest”. In: *Motivation and emotion* 12.4, pp. 309–331.

de Gemmis, Marco, Pasquale Lops, and Marco Polignano (2017). “Recommender Systems, Basics Of”. In: ed. by Reda Alhajj and Jon Rokne, pp. 1–13. DOI: 10.1007/978-1-4614-7163-9_110158-1.

de Gemmis, Marco, Pasquale Lops, and Giovanni Semeraro (2016). “Emotion Detection Techniques for the Evaluation of Serendipitous Recommendations”. In: ed. by Marko Tkalčič, Berardina De Carolis, Marco de Gemmis, Ante Odić, and Andrej Košir, pp. 357–376. DOI: 10.1007/978-3-319-31413-6_17.

de Gemmis, Marco, Pasquale Lops, Giovanni Semeraro, and Cataldo Musto (2015a). “An Investigation on the Serendipity Problem in Recommender Systems”. In: *Information Processing & Management* 51.5, pp. 695–717. DOI: 10.1016/j.ipm.2015.06.008.

de Gemmis, Marco, Pasquale Lops, Cataldo Musto, Fedelucio Narducci, and Giovanni Semeraro (2015b). “Semantics-Aware Content-Based Recommender Systems”. In: ed. by Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, pp. 119–159. DOI: 10.1007/978-1-4899-7637-6_4.

Deng, Shuiguang, Dongjing Wang, Xitong Li, and Guandong Xu (2015). “Exploring user emotion in microblogs for music recommendation”. In: *Expert Systems with Applications* 42.23, pp. 9284–9293.

Ekman, Paul, Wallace V Friesen, Maureen O’Sullivan, Anthony Chan, Irene Diacoyanni-Tarlatzis, Karl Heider, Rainer Krause, William Ayhan LeCompte, Tom Pitcairn, Pio E Ricci-Bitti, et al. (1987). “Universals and cultural differences in the judgments of facial expressions of emotion.” In: *Journal of Personality and Social Psychology* 53.4, p. 712.

Ferwerda, Bruce, Emily Yang, Markus Schedl, and Marko Tkalcic (2015). “Personality Traits Predict Music Taxonomy Preferences”. In: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA ’15. Seoul, Republic of Korea: ACM, pp. 2241–2246. ISBN: 978-1-4503-3146-3. DOI: 10.1145/2702613.2732754.

Fiori, M., A. Lintas, S. Mesrobian, and A. E. P. Villa (2013). “Effect of Emotion and Personality on Deviation from Purely Rational Decision-Making”. In: *Decision Making and Imperfection*. Vol. 474. Studies in Computational Intelligence. Springer, pp. 129–161. DOI: 10.1007/978-3-642-36406-8_5.

Frijda, Nico H and Batja Mesquita (1994). “The social roles and functions of emotions.” In: *Emotion and culture: Empirical studies of mutual influence*, pp. 51–87.

Goddard, Michelle (2017). “The EU General Data Protection Regulation (GDPR): European regulation that has a global impact”. In: *International Journal of Market Research* 59.6, pp. 703–705.

González, Gustavo, Beatriz López, and Josep Lluís de la Rosa (2004). “Managing Emotions in Smart User Models for Recommender Systems”. In: *ICEIS 2004, Proceedings of the 6th International Conference on Enterprise Information Systems, Porto, Portugal, April 14-17, 2004*, pp. 187–194.

Gonzalez, Gustavo, Josep Lluis de la Rosa, Miquel Montaner, and Sonia Delfin (2007). “Embedding Emotional Context in Recommender Systems”. In: *Proceedings of the 23rd International Conference on Data Engineering Workshops, ICDE 2007, 15-20 April 2007, Istanbul, Turkey.* IEEE Computer Society, pp. 845–852. DOI: 10.1109/ICDEW.2007.4401075.

Han, Byeong-jun, Seungmin Rho, Sanghoon Jun, and Eenjun Hwang (2010). “Music emotion classification and context-based music recommendation”. In: *Multimedia Tools and Applications* 47.3, pp. 433–460.

Jameson, Anthony, Martijn C. Willemsen, Alexander Felfernig, Marco de Gemmis, Pasquale Lops, Giovanni Semeraro, and Li Chen (2015). “Human Decision Making and Recommender Systems”. In: *Recommender Systems Handbook*. Ed. by Francesco Ricci, Lior Rokach, and Bracha Shapira. Springer, pp. 611–648. DOI: 10.1007/978-1-4899-7637-6_18.

Jaro, Matthew A (1989). “Advances in Record-linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida”. In: *Journal of the American Statistical Association* 84.406, pp. 414–420. DOI: 10.1080/01621459.1989.10478785.

Joho, Hideo, Jacopo Staiano, Nicu Sebe, and Joemon M. Jose (2011). “Looking at the Viewer: Analysing Facial Activity to Detect Personal Highlights of Multimedia Contents”. In: *Multimedia Tools and Applications* 51.2, pp. 505–523. DOI: 10.1007/s11042-010-0632-x.

Kaminskas, Marius and Francesco Ricci (2016). "Emotion-Based Matching of Music to Places". In: *Emotions and Personality in Personalized Services: Models, Evaluation and Applications*. Ed. by Marko Tkalčič, Berardina De Carolis, Marco de Gemmis, Ante Odić, and Andrej Košir. Cham: Springer International Publishing, pp. 287–310. ISBN: 978-3-319-31413-6. DOI: 10.1007/978-3-319-31413-6_14.

Kazai, Gabriella, Iskander Yusof, and Daoud Clarke (2016). "Personalised News and Blog Recommendations based on User Location, Facebook and Twitter User Profiling". In: *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, pp. 1129–1132. DOI: 10.1145/2911451.2911464.

Kim, Jonghwa and Elisabeth Andre (2008). "Emotion recognition based on physiological changes in music listening". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.12, pp. 2067–2083. DOI: 10.1109/TPAMI.2008.26.

Kim, Youngmoo E., Erik M. Schmidt, Raymond Migneco, On G. Morton, Patrick Richardson, Jeffrey Scott, Jacquelin A. Speck, and Douglas Turnbull (2010). "Music Emotion Recognition: a State of the Art Review". In: *Proceedings of the 11th International Society for Music Information and Retrieval Conference*, pp. 255–266.

Klein, G. (2015). "A naturalistic decision making perspective on studying intuitive decision making". In: *Journal of Applied Research in Memory and Cognition* 4.3, pp. 164–168. DOI: 10.1016/j.jarmac.2015.07.001.

Klein, Gary A, Roberta Calderwood, and Anne Clinton-Cirocco (2010). "Rapid Decision Making on the Fire Ground: The Original Study Plus a Postscript". In: *Journal of Cognitive Engineering and Decision Making* 4.3, pp. 186–209. DOI: 10.1518/155534310X12844000801203.

Kosinski, Michal, Sandra C Matz, Samuel D Gosling, Vesselin Popov, and David Stillwell (2015). "Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines". In: *American Psychologist* 70.6, p. 543. DOI: 10.1037/a0039210.

Lee, Seungjae, Jung Hyun Kim, Sung Min Kim, and Won Young Yoo (2011). "Smoodi: Mood-based music recommendation player". In: *Multimedia and Expo (ICME), IEEE International Conference on*. IEEE, pp. 1–4. DOI: 10.1109/ICME.2011.6012116.

Lee, Su-In, Honglak Lee, Pieter Abbeel, and Andrew Y. Ng (2006). "Efficient L1 Regularized Logistic Regression". In: *Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, July 16-20, 2006, Boston, Massachusetts, USA*, pp. 401–408.

Li, Lei, Li Zheng, Fan Yang, and Tao Li (2014). "Modeling and broadening temporal user interest in personalized news recommendation". In: *Expert Systems with Applications* 41.7, pp. 3168–3177.

Liang, Weiming, Haoran Xie, Yanghui Rao, Raymond YK Lau, and Fu Lee Wang (2018). "Universal affective model for Readers' emotion classification over short texts". In: *Expert Systems with Applications* 114, pp. 322–333.

Loewenstein, G. and J. S. Lerner (2003a). "The role of affect in decision making". In: *Handbook of Affective Science*. Ed. by R. Davidson, H. Goldsmith, and K. Scherer. Oxford University Press, pp. 619–642.

Loewenstein, George and Jennifer S. Lerner (2003b). "The Role of Affect in Decision Making". In: *Handbook of Affective Science* 619.642.

Ma, Chao, Chen Zhu, Yanjie Fu, Hengshu Zhu, Guiquan Liu, and Enhong Chen (2017). "Social User Profiling: A Social-Aware Topic Modeling Perspective". In: *International Conference on Database Systems for Advanced Applications*. Springer, pp. 610–622. DOI: 10.1007/978-3-319-55699-4_38.

Malhotra, Anshu, Luam Totti, Wagner Meira Jr, Ponnurangam Kumaraguru, and Virgilio Almeida (2012). "Studying user footprints in different online social networks". In: *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*. IEEE Computer Society, pp. 1065–1070. DOI: 10.1109/ASONAM.2012.184.

Medhat, Walaa, Ahmed Hassan, and Hoda Korashy (2014). "Sentiment analysis algorithms and applications: A survey". In: *Ain Shams Engineering Journal* 5.4, pp. 1093–1113. DOI: 10.1016/j.asej.2014.04.011.

Mohammad, Saif M. and Svetlana Kiritchenko (2015). "Using Hashtags to Capture Fine Emotion Categories from Tweets". In: *Computational Intelligence* 31.2, pp. 301–326. ISSN: 1467-8640. DOI: 10.1111/coin.12024.

Papadakis, Harris, Costas Panagiotakis, and Paraskevi Fragopoulou (2017). "SCoR: a synthetic coordinate based recommender system". In: *Expert Systems with Applications* 79, pp. 8–19.

Park, Han-Saem, Ji-Oh Yoo, and Sung-Bae Cho (2006). "A context-aware music recommendation system using fuzzy bayesian networks with utility theory". In: *International Conference on Fuzzy Systems and Knowledge Discovery*. Springer, pp. 970–979. DOI: 10.1007/11881599_121.

Peters, E. (2006). "The Functions of Affect in the Construction of Preferences". In: *The Construction of Preference*. Ed. by S. Lichtenstein and P. Slovic, pp. 454–463.

Pfister, H.-R. and G. Böhm (1992). "The function of concrete emotions in rational decision making". In: *Acta Psychologica* 80, pp. 199–211.

Pfister, H.-R. and G. Böhm (2008). "The multiplicity of emotions: A framework of emotional functions in decision making". In: *Judgment and Decision Making* 3.1, pp. 5–17.

Picard, Rosalind W. (1997). *Affective Computing*. Cambridge, MA, USA: MIT Press. ISBN: 0-262-16170-2.

Picard, Rosalind W, Seymour Papert, Walter Bender, Bruce Blumberg, Cynthia Breazeal, David Cavallo, Tod Machover, Mitchel Resnick, Deb Roy, and Carol Strohecker (2004). "Affective learning - A manifesto". In: *BT Technology Journal* 22.4, pp. 253–269. DOI: 10.1023/B:BTTJ.0000047603.37042.33.

Poirson, Emilie and Catherine Da Cunha (2018). "A recommender approach based on customer emotions". In: *Expert Systems with Applications*.

Polignano, Marco, Marco de Gemmis, Fedelucio Narducci, and Giovanni Semeraro (2017a). “Do You Feel Blue? Detection of Negative Feeling from Social Media”. In: *Conference of the Italian Association for Artificial Intelligence*. Springer, pp. 321–333. DOI: 10.1007/978-3-319-70169-1_24.

Polignano, Marco, Pierpaolo Basile, Gaetano Rossiello, Marco de Gemmis, and Giovanni Semeraro (2017b). “Learning inclination to empathy from social media footprints”. In: *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. ACM, pp. 383–384. DOI: 10.1145/3079628.3079639.

— (2017c). “User’s Social Media Profile As Predictor of Empathy”. In: *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. UMAP ’17. Bratislava, Slovakia: ACM, pp. 386–390. ISBN: 978-1-4503-5067-9. DOI: 10.1145/3099023.3099103.

Qian, Yongfeng, Yin Zhang, Xiao Ma, Han Yu, and Limei Peng (2019). “EARS: Emotion-aware recommender system based on hybrid information fusion”. In: *Information Fusion* 46, pp. 141–146.

Ryan, Tracii and Sophia Xenos (2011). “Who uses Facebook? An Investigation into the Relationship between the Big Five, Shyness, Narcissism, Loneliness, and Facebook Usage”. In: *Computers in Human Behavior* 27.5, pp. 1658–1664. DOI: 10.1016/j.chb.2011.02.004.

Sarwar, Badrul, George Karypis, Joseph Konstan, and John Riedl (2001). “Item-based Collaborative Filtering Recommendation Algorithms”. In: *Proceedings of the 10th International Conference on World Wide Web*. WWW ’01. Hong Kong, Hong Kong: ACM, pp. 285–295. ISBN: 1-58113-348-0. DOI: 10.1145/371920.372071.

Schilit, Bill, Norman Adams, and Roy Want (1994). “Context-aware computing applications”. In: *Mobile Computing Systems and Applications, 1994. WMCSA. First Workshop on*. IEEE, pp. 85–90.

Schwarz, Norbert (2000). “Emotion, cognition, and decision making”. In: *Cognition & Emotion* 14.4, pp. 433–440.

Shanahan, James G., Yan Qu, and Janyce Wiebe, eds. (2006). *Computing Attitude and Affect in Text: Theory and Applications*. Vol. 20. The Information Retrieval Series. Springer. ISBN: 978-1-4020-4026-9. DOI: 10.1007/1-4020-4102-0.

Shao, Zongru, Rajarathnam Chandramouli, KP Subbalakshmi, and Constantine T Boyadjiev (2019). “An analytical system for user emotion extraction, mental state modeling, and rating”. In: *Expert Systems with Applications* 124, pp. 82–96.

Shi, Yue, Martha Larson, and Alan Hanjalic (2010). “Mining Mood-specific Movie Similarity with Matrix Factorization for Context-aware Recommendation”. In: *Proceedings of the Workshop on Context-Aware Movie Recommendation*. CAMRa ’10. Barcelona, Spain: ACM, pp. 34–40. ISBN: 978-1-4503-0258-6. DOI: 10.1145/1869652.1869658.

Soleymani, M. and M. Pantic (2012). “Human-centered Implicit Tagging: Overview and Perspectives”. In: *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, pp. 3304–3309. DOI: 10.1109/ICSMC.2012.6378301.

Soleymani, Mohammad, Maja Pantic, and Thierry Pun (2012). “Multimodal Emotion Recognition in Response to Videos”. In: *IEEE Transaction on Affective Computing* 3.2, pp. 211–223. DOI: http://doi.ieeecomputersociety.org/10.1109/T-AFFC.2011.37.

Tkalcic, Marko, Urban Burnik, and Andrej Kosir (2010). “Using Affective Parameters in a Content-based Recommender System for Images”. In: *User Modeling User-Adapted Interaction* 20.4, pp. 279–311. DOI: 10.1007/s11257-010-9079-z.

Tkalcic, Marko, Urban Burnik, Ante Odic, Andrej Kosir, and Jurij F. Tasic (2012). “Emotion-Aware Recommender Systems - A Framework and a Case Study”. In: *ICT Innovations 2012 - Secure and Intelligent Systems, Ohrid, Macedonia, 12-15 September, 2012*, pp. 141–150.

Tkalcic, Marko, Ante Odic, Andrej Kosir, and Jurij F. Tasic (2013). “Affective Labeling in a Content-Based Recommender System for Images”. In: *IEEE*

Transactions on Multimedia 15.2, pp. 391–400. DOI: 10.1109/TMM.2012.2229970.

Tran, Thanh, Kyumin Lee, Yiming Liao, and Dongwon Lee (2018). "Regularizing matrix factorization with user and item embeddings for recommendation". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 687–696.

Winoto, Pinata and Tiffany Y Tang (2010). "The role of user mood in movie recommendations". In: *Expert Systems with Applications* 37.8, pp. 6086–6092. DOI: 10.1016/j.eswa.2010.02.117.

Yang, Jaewon and Jure Leskovec (2011). "Patterns of temporal variation in online media". In: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*. ACM, pp. 177–186. DOI: 10.1145/1935826.1935863.

Yang, Yi-Hsuan, Ya-Fan Su, Yu-Ching Lin, and Homer H. Chen (2007). *Music emotion recognition: the role of individuality*, pp. 13–22. DOI: 10.1145/1290128.1290132.

Zhang, Lei and Bing Liu (2017). "Sentiment Analysis and Opinion Mining". In: *Encyclopedia of Machine Learning and Data Mining*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, pp. 1152–1161. ISBN: 978-1-4899-7687-1. DOI: 10.1007/978-1-4899-7687-1_907. URL: https://doi.org/10.1007/978-1-4899-7687-1_907.

Zheng, Y., B. Mobasher, and R. D. Burke (2013). "The Role of Emotions in Context-aware Recommendation". In: *Proceedings of the 3rd Workshop on Human Decision Making in Recommender Systems, in conjunction with the 7th ACM Conference on Recommender Systems (RecSys 2013)*. Ed. by L. Chen, M. de Gemmis, A. Felfernig, P. Lops, F. Ricci, G. Semeraro, and M. C. Willemsen. Vol. 1050. CEUR Workshop Proceedings. CEUR-WS.org, pp. 21–28.

Zheng, Yong, Bamshad Mobasher, and Robin Burke (2016). "Emotions in Context-Aware Recommender Systems". In: *Emotions and Personality in Personalized Services: Models, Evaluation and Applications*. Ed. by Marko Tkalčič, Berardina De Carolis, Marco de Gemmis, Ante Odić, and Andrej Košir. Springer International, pp. 311–326. ISBN: 978-3-319-31413-6. DOI: 10.1007/978-3-319-31413-6_15.

Appendix A. Integrative Tables

Table A.10: List of Symbols used in equations

Symbol	**Description**
I	Set of items
i_j	j-esim item
u	Active user
r_j	Rating left by the user u at the item i_j
$p_{t,j}$	Preference of the user u for the item i_j at time t
P_u	Set of user preferences $p_{t,j}$
t	Time value
A_u	Set of affective states of the user u
$\vec{s_t}$	Affective state of the user u at time t
e_i	Activation score for the emotion e
$\vec{a_j}$	Affective state associated with the item i_j
AP_u	Set of tuple affective states - preferences
i_z	Item to be recommended
α	Weight used for changing the importance of the affective similarity in the recommendation score
β	Weight used for changing the importance of the item similarity in the recommendation score
W	Time window for collecting messages on SMSs
M	Set of user messages
m_i	i-esim user message
γ_i	decay factor

Table A.11: List of lyrics websites used for crawling songs textual content

	URL
Lyrics Websites	azlyric.com, elyrics.com, lyricsera.com, sweetlyrics.com, lyricsfreak.com, songlyrics.com, lyricsmode.com, metrolyrics.com, flashlyric.com, lololyrics.com, songtexte.com, lyricsmania.com, songmeanings.com, lyricofsong.com