



HAL
open science

A recommender approach based on customer emotions

Emilie Poirson, Catherine Da Cunha

► **To cite this version:**

Emilie Poirson, Catherine Da Cunha. A recommender approach based on customer emotions. *Expert Systems with Applications*, 2019, 122, pp.281-288. 10.1016/j.eswa.2018.12.035 . hal-01987342

HAL Id: hal-01987342

<https://hal.science/hal-01987342v1>

Submitted on 21 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

A recommender approach based on customer emotions

Emilie Poirson^a and Catherine Da Cunha^b

^a Ecole Centrale de Nantes, LS2N (Laboratoire des Sciences du Numerique de Nantes)
UMR CNRS 6004, 1, rue de la Noe, 44300 Nantes, France
emilie.poirson@ec-nantes.fr, +33 2 40 37 69 57
Corresponding author.

^b Ecole Centrale de Nantes, LS2N (Laboratoire des Sciences du Numerique de Nantes)
UMR CNRS 6004, 1, rue de la Noe, 44300 Nantes, France
catherine.da-cunha@ec-nantes.fr, +33 2 40 37 69 57

Point-to-point responses to the reviewers

The authors would firstly like to thank the reviewers for their constructive remarks. They truly helped the authors to clarify their choices, to express their ideas precisely and to highlight the work done. The authors would like to apologize for the linguistic errors. The paper has now been proof read by a native English speaker for more fluent reading.

Reviewer #1

This paper proposes to take into account the users' emotions (or the similarity between a pair of users) for the Recommender Systems. Although such an idea is interesting in some degree, the proposed approach is rather straightforward from a methodology point of view and difficult to use in applications due to the unavailable emotion information. From a methodology point, the proposed approach is just including the emotion factors as the additional attributes in the users' similarity calculation, and there is little real novelty from a methodology point of view. From an application point of view, it is very difficult to get the emotion information from users, except using the questionnaires as did in this proposed study. Unfortunately it is infeasible to use questionnaires to collect users' emotion information given a large number of users and the required effort from users to answer such questionnaires.

To select the appropriate emotions for our product, we based our work on the Differential Emotions Scale (DES) (Izard 1977), and adaptations of the DES (Philippot 1993; Gross and Levenson 1995). This proposal includes 8 emotions: Amused, Sad, Angry, Scared, Attentive, Anxious, Surprise and Happy. In a previous study, we showed that the 8 emotions are of equal relevance (Poirson, Da Cunha, and Petiot 2014).

The reviewer is right when he/she says that the tools of representation or collection of emotions are not very innovative. We do not claim a significant contribution on this subject. The purpose of this paper is to evaluate our hypothesis of a link between the emotions and the general opinion on a product. Indeed, we did not really focus on collecting emotions, which is a research subject in its own. In a broader study, it could be interesting to work on an adaptable collection method depending on the specific product studied. The questionnaires are indeed time-consuming and not adapted to extensive studies. More and more new techniques are being developed to automatically capture these emotions, through semantic analysis of user reviews or facial detections. In our study and to be relevant, we wanted to focus on smaller groups for which we could actually and simply collect the information rather than simulate it.

To be clearer in the paper, the introduction and the research question are re-written in the current version of the paper.

Further for the cultural products such as movies, where the proposed approach is applicable, the users' preferences in fact include the emotion factors and evaluations but just implicitly. But the Recommender Systems based on the preference similarity between users gives a much simpler and more effective approach as the preference similarity includes implicitly much more comprehensive information such as emotion, culture, style, taste etc and such information is widely available.

Absolutely. "Preferences" aggregate a lot of information on culture, emotion, experiment. Our purpose is not to simply add columns of additional information to improve the existing model. The aim is to check if the actual "preference" is really the best indicator for recommendations. Emotions could be associated with the preference, but could perhaps also replace this global evaluation. The implicit information of the preference could be found in the emotions. That is what we want to check.

We understood that the presentation of the study was not clear enough for the reader. The abstract, introduction and context of the study have been completely re-written with a better positioning of the study in the context (section 1 and 2, new subsection 2.2).

Reviewer #2

1 - Scenario

Description of the work that does not require answer.

2 - Presentation: The manuscript has not been carefully written, since there are many typos and a number of sentences should be better expressed, also because in many cases there is an inappropriate use of punctuation marks. Some examples can be already found in the "Abstract", where sentences such as "The performance criterion computed, measuring the relevance of the proposed products the recommender is then applied to the second part of the study: a test with real data." or "The efficiency of our method is studied, depending on the sparsity of this database and other parameters of the model." should be corrected and rewritten in a clearer way, as well as many other sentences in the paper (e.g., "We choose not to weight the correlation thanks to the number common products." in Section 3.2 or "Further research will focus on a methodology to systematize the acquisition of emotional descriptors" in Section 6).

Certain sentences were indeed difficult to understand. They have been reformulated with attention to punctuation.

Another problem is related to the writing contractions that should be absolutely avoided in a formal work such as a scientific paper, e.g.: "A good evaluation doesn't mean that the product is good." and "a family won't take into account the same arguments" in Section 2.2; "That's the reason why" and "Let's underline the difficulty" in Section 3.1; "Let's focus on the number of products" in Section 3.2;

The contractions have been replaced.

The nomenclature should be coherent and the authors should adopt a more technical language. For instance, in different parts of the document the authors use the terms "customers", "consumers", "subjects", and "users" to refer to the same entity, and non-technical terms such as "notes on preferences" (instead of "preferences data" or other similar terms).

The terms are now homogeneous. The product in our application case is "film". In the experimentation (4th section), we speak about "Subjects". "Subject" is the correct terminology in perceptive evaluation. It allows us to distinguish the differences between simulated data (customers) and experimental data produced by subjects.

We have also removed non-technical formulations.

The typography errors listed (and others) have been corrected.

Also in terms of in terms of writing accuracy the paper shows many issues. In addition to the aforementioned typos, there are sentences such as "marketing, engineering, economics and engineering" in Section 2.3, or "The approach is described in Figure 2. The proposed approach relies on our research hypotheses:" instead of "The proposed approach, described in Figure 2, relies on our research hypotheses:", and so on.

The paper has been proofread by a native English speaker. We hope that the work provided adds significant value to the reading comprehension and fluency of the paper.

3 - Organization: About the organization of the paper contents, some elements go out of page column border and should be appropriately resized (e.g., Table 1 and Table 2).

The corrections have been made.

Some efforts should be also made to improve the general organization of the paper, since the introductory part is a slightly confused. I suggest to the authors to organize the Section 2 in two parts, one devoted to the background aspects, where they underline the main research problems to be handled, and one devoted to the related work. A separate section should be used to introduce the idea behind their work and its related scenario.

We understand the confusion. The second section is now in 2 parts: the state of the art and the work done in this context where we present our hypotheses linked to the bibliography.

4 - Implementation: Also the implementation section presents some problems related to the way in which the authors present their approach. A bulleted or numbered list, where all the needed steps are summarizes before their detailed explanation, should be placed at the beginning of the section 3.

Section 3 has been restructured. The global presentation of the approach is now set out at the beginning of section 3.2.

5 - Experiments: The Section 4 starts with the sentence "The computational experimentation was lead using Python 3.1 programming. The velocity of the computation was not evaluated, since for the real application the calculation of the recommender system can be done off-line.". Such sentence needs to be rewritten in a clearer way, for instance:

*(1) what it means "computational experimentation"? Maybe they meant "the experiments";
(2) what it means "was lead using"? Maybe they meant "was performed by using";
(3) what it means "The velocity of the computation"? Maybe they meant "computational load". The authors should be use a more technical language.*

The added-value of Section 4.1 (Simulated case) is questionable, especially for the small number of the involved subjects (more appropriately, "users") and films (more appropriately, "items").

A new subsection (environment) aimed to provide all the information about the experimental environment (used computers and type, programming language, etc.) must be added as first subsection of the Experimentation section.

A section describing the experimental protocol was added at the beginning of section 4. It stresses the objectives of implementations on both simulated and real data.

Proofreading by a native English reviewer has helped to remove unclear sentences.

The objectives of the simulated case are highlighted: this controlled environment, verifying assumption H1, enables us to verify assumptions H2 and H3. It also enables us to tune the parameter of our algorithms (threshold) and verify its resistance to noise and sparsity.

6 - Results and conclusion: The "Result and discussion" section should be reorganized in order to better present the information (maybe by recurring to further subsections and/or bulleted or numbered lists) and, in any case, all the obtained results (simulated and real case) must be clearly summarized at the end of the section .

The "Conclusion" section should be only focused on the description of the contributions given by the proposed work, avoiding to involve information already given in other parts of the paper.

Bullet lists have been added in order to highlight the obtained results. The conclusion has been rewritten in order to avoid redundancies with previous sections. We hope the new version stresses the contributions of the paper.

7 - References: About the references, I noted that almost all of them are very dated: 10 of them have been published from 1977 to 1999 and only 6 out of 30 are quite recent (i.e.,

published in the last 5 years).

Furthermore, there are some references do not have a direct connection with the research area taken into account in the paper. Some examples are the following: (1) Nakada, K. 1997.

"Kansei engineering research on the design of construction machinery." *International Journal of Industrial Ergonomics* 19: 129-146.;

(2) Pazzani, M. 1999. "A framework for collaborative, content-based and demographic filtering." *Artificial Intelligence Review* 13 (5-6): 393-408.;

(3) Yanagisawa, Hideyoshi. 2011. *Kansei Quality in Product Design*, 289-310. London: Springer London.

We understand this remark. It is difficult to speak about emotion in product design without citing Hideyoshi Yanagisawa, for example, on the topic of Kansei. We realized that the paper is not sufficiently "product design oriented" to include the work of Kansei in the state of the art, the connection not being direct. We have deleted the 3 papers from the bibliography.

Finally, some significant works focused on the context taken into account by the authors are missing and should be added, such as, for instance:

1 "News recommendation via hypergraph learning: encapsulation of user behavior and news content." Li, Lei, and Tao Li. "Proceedings of the sixth ACM international conference on Web search and data mining, 2013";

2 "Semantic Coherence-based User Profile Modeling in the Recommender Systems Context. R Saia, L Boratto, S Carta - *KDIR*, 2014";

3 "Modeling and broadening temporal user interest in personalized news recommendation. Li, Lei, et al. - *Expert Systems with Applications*, 2014;

4 "A semantic approach to remove incoherent items from a user profile and improve the accuracy of a recommender system. R Saia, L Boratto, S Carta - *Journal of Intelligent Information Systems*, 2016".

The authors thank the reviewer for this advice. The 4 papers cited indeed bring information and example in this scope with, for example, application to film. We have added them to the bibliography.

8 - Final observations: The major weakness of this paper is mainly determined by the union of two aspects, one formal and one substantial.

The formal aspect concerns the lack of care in its writing and the numerous grammatical and semantic errors, as well as other minor issues, such as the inconsistency in the nomenclature, the writing contractions, and the use of an inappropriate and non-technical terminology.

On the light of the aforementioned observations, I would suggest the authors to carefully proof-read the whole paper, involving in this process a native English speaker.

Completely right. We have followed this advice.

About the substantial aspect, the authors' choice to use for the validation of their approach a synthetic-data and an unverifiable real-data is scientifically questionable. They must better motivate this choice throughout the paper.

Although the scientific contribution can be identified in the work done by the authors, it is necessary that they present it in a better way, emphasizing the advantages associated with the proposed approach.

We have made major revisions to the paper, paying a particular attention to the enhancement of the results and the precision of the methodology.

A final note concerns the references to literature, since they should be reviewed, by removing some that are non-relevant for the context taken into account, adding other ones more

pertinent and updated.

The bibliography has been updated.

In conclusion, I consider that the proposed paper is interesting and it fits within the scope of the "Expert Systems With Applications" journal, but before the publication it needs to be improved through a number of corrections and improvements indicated in my review.

Reviewer #3

The paper makes an interesting proposal that consider both emotions and ratings of users and hybridates a similarity measure using each of them to combine it with a multiplication to obtain an aggregate similarity that considers both in the recommendation process. A well-supported experimentation is performed to validate the proposal and test the hypotheses. The paper is justified and the presentation of the ideas is clear. However, there are some issues that need to be addressed: The passive voice is overused in the abstract, which makes it difficult to understand it.

The abstract as been reworded a lot, paying particular attention to the voice used.

One of the most important parts of the paper has a typo in it: the paragraph of the introduction that describes the main contributions of the paper. "The main contributions of this paper are: (1) we propose hypothesis on the link between emotions, preferences and recommendation and we validate them; (2) we propose modelisation of a customer panel (by profile of customers) and an algorithm to improve the quality of the recommendations; (3) we evaluate the performance of the proposed algorithm on simulated data and a real application case on movies".

Please, revise the whole paper thoroughly to improve the quality of english removing unneceSsary use of passive voice, and checking for typos. Some of them are shown here: - futur is not a word in English. Neither 'common' is. - 'recommendation' -¿ recommendation - "This first part give a matrix of similarity in terms of preferences" -¿ gives - "the proposition of a method that improve the recommendation" -¿ improves - recommander -¿ recommender

The paper has been proofread by a native English speaker. We hope that the work provided adds significant value to the reading comprehension and fluency of the paper.

In the reference "Konstan, J.A., and J.Redl. 2012. "Recommender Systems: From Algorithms to User Experience." User Modeling and User-Adapted Interaction 22: 101-123", the surname of the second author is misspelled.

The correction has been made.

A recommender approach based on customer emotions

E. Poirson^a and C. Da Cunha^a

^a Ecole Centrale de Nantes, LS2N (Laboratoire des Sciences du Numerique de Nantes) UMR CNRS 6004, France

ARTICLE HISTORY

Compiled January 16, 2018

ABSTRACT

On electronic commerce websites, the recommender is typically based on a global evaluation of the product by the customer. This preference information is interesting, but not sufficient. Our approach is based on the assumption of the inevitable role of emotions in product appreciation. Our purpose is to evaluate the performance of the approach based on emotions and to compare it with the traditional approach of unique preference. The first step is to collect data. In the first part of the study, we simulate a database of customer/product/evaluation to test our method and to study the influence of matrix sparsity. The second part of the study is applied to real data. The product considered to illustrate the approach is film, a product for which preferences are mostly guided by emotions. We collect the data via an online questionnaire, given a sparse matrix of evaluation (people have not seen all the films). A use case, which consist in making spontaneous recommendations to customers on the basis of their previous ratings, is described. Considering that the behavior of a customer is not random, the notion of user profiles in terms of preference and emotions is introduced. The proposed recommender is tested by removing preference data and predicting it based on the rest of the database. A performance criterion is calculated, measuring the relevance of the proposed products. The results show that (1) the matrix sparsity does not have a significant influence on the results and robustness of our method and (2) consideration of customers' emotional assessments of products improves recommendation performance.

KEYWORDS

Recommender Systems, Preference, Emotions, Customer profile, User evaluations.

1. Introduction

As electronic commerce websites offer a huge number of products and services for sale, choosing the right good is sometimes complicated for a website visitor. These websites often use recommender systems to help potential buyers to find products or services that best meet their needs and preferences. Basic recommenders focus on customers experience. The costumer is identified by his/her purchase history (products and dates) and objective data such as geographical information or age for example ; representing a huge amount of data. To enhance the product feedback databases, costumers are now asked to give their opinion on websites. For example, on several influential online international market websites such as Amazon, ebay or Aliexpress, the user is invited to rate products on 5 grades (stars) or to click on a 'thumb up' or 'thumb down'. In March 2017, for example, more than 350 million

CONTACT E. Poirson Email: emilie.poirson@ec-nantes.fr

CONTACT E. Poirson Email: emilie.poirson@ec-nantes.fr

unique visitors to Amazon, ebay or wallmart were therefore asked for their opinions ¹. Big data tools are, therefore, required to manage this data for product recommendation. Our study focuses on recommenders, to help users navigate the ever-increasing number of offers that meet their expectations.

Three types of data are collected in such activities. Firstly, objective data on products and customers is recorded, once. Secondly, hedonic data, expressing like or dislike of a product. And lastly, perceptive or emotional data, describing the product individually, based on customer experience. The work done aims to find a link between the data types, particularly between hedonic and perceptive data in order to improve recommendation quality. In this paper, we will use the term "preference" - even if it is sometimes more adapted to relative data - instead of 'hedonic data' to consider the global evaluation of a product. The notion of 'preference profiles' for a customer is also introduced.

The second section of the paper starts by describing the state of the art in filtering techniques including clusterization and emotion in design, underlining the research problem to handle. We will then explain our contribution in detail. The third section describes the methodology and the fourth focuses on the experimentation of a real case. This is followed by the results and discussion section.

The main contributions of this paper are:

- (1) a proposal and confirmation of the link between emotions, preferences and recommendation;
- (2) a proposal for modeling a customer panel (by customers profile) and an algorithm to improve the recommendation quality;
- (3) an evaluation of the performance of the proposed algorithm on simulated data and a real application case on films.

2. Context of the study

2.1. State of the art

Recommendation systems

Recommender systems are based on a variety of filtering techniques. The most popular is collaborative filtering (CF), which provides a recommendation list from the user profile and contextual parameters, associated with community data (Herlocker, J.Konstan, and J.Riedl 2000; Jannach and Kreutler 2007; Konstan and J.Riedl 2012). Content-based filtering focuses on product features as described on a webpage (Degemmis, Lops, and Semeraro 2007). Today, personalized information recommenders exist, based on the hybridization of collaborative and content-based techniques (Serrano-Guerrero et al. 2011; Porcel et al. 2012). The same author as your previous book, the product that other people bought at the same time as the one you are interested in, a product highly ranked by people who give the same opinion as you in other situations, a product liked by someone of the same age... as many potential purchases, in a large number of applications. But how far is this from your real desire or need? Filtering on purely objective data appears limited, even if some parameters are adapted for certain types of products. In (Li et al. 2014), time is considered in the browsing of a news article, highlighting the difference between immediate interest (potential frequent evolution) and long term interest (more stable). By the use of hypergraph (Li and Li 2013), data is correlated and recommendations are made based on the 'time' parameter without the intervention of the user to explain his choices. The system can suffer from wide inaccuracies implying poor 'quality of recommendations for the users' (Sarwar et al. 2000). One of the limits of profiling users on objective data is that distances between numerical attributes (age)

¹<https://www.statista.com/statistics/271450/monthly-unique-visitors-to-us-retail-websites/>

and category attributes (work category) are calculated. In order to aggregate this data and to define distances between users, several models are proposed (Al-Shamri 2016). Al-Shamri underlines that age must be processed as a fuzzy variable, confirming the difficulty to classify people according to purely objective data. Reducing people to an age or socio-professional category would be caricatural and contrary to product innovation and universality. There is, thus, no good method to recommend a product basing the decision only on objective data. The second limitation is the wrong interpretation that one can make of the product evaluations resulting from a content-based filtering. Working on the product itself does not make sense. The evaluation received on the product does not reflect the quality of the product (in terms of conformity), but rather the satisfaction of the user through his/her interaction with the product (Englern, Winter, and Schulz 2015). A good evaluation does not mean that the product is good. It means that the product fits the needs of this customer, what is more, in his current mood (Winoto and Tang 2010). It, therefore, seems to us essential to take into account information concerning the interactions of these two elements: the emotional state of the user according to the use of the product. (Deng et al. 2015) defined, for example, the links between the use of a product (in this case, choice of song) and its environment that is here the emotional context. The aim was to find the explanation of a musical choice through the emotional characteristics of the user. The real interest, in the case of a recommender, is not whether the user liked the product, but why he liked it. For example, in the tourism industry (Jannach, Zanker, and Fuchs 2014), a family will not take into account the same arguments as a young solo-traveler to evaluate a hotel in an identical way. A product is evaluated in a context and in a precise emotional state.

User emotions

Since the sixties, many disciplines such as marketing, economics and engineering have taken into account user perceptions and emotions, as emotions influence purchase decisions. Perception is the identification and interpretation of an identical way information via human sensory organs and nervous systems. Emotion is the affective reaction that manifests itself through physical disturbances. Emotions can be viewed as mechanisms that return signals in response to favorable or harmful events (Desmet, Porcelijn, and van Dijk 2007). In application cases involving culture or fashion, emotion particularly influences our choices and our decision-making. We thus focus in this study on "emotional attributes"; terms related to the affective state of the customer pleasure and displeasure. Research in customer behavior has discovered, and subsequently confirmed the importance and the dominance of emotions and affective states in decision making (Richins 1997). Several studies have shown that taking into account the preference of a customer begins with an understanding of the product's emotional perception to use them in recommender(Tkalčič, Kosir, and Tasic 2011). Differences between customers' preferences may be due to the fact that they perceive the same product differently (Shang et al. 2000). (Lim and Kim 2017) also acknowledge the close relationship between emotions and overall preferences (evaluated through ratings). They chose to use tag analysis to highlight and monitor this link. Their method enabled to improve future recommendation. One of the limitations of their work is the use of an implicit expression of the emotions (tag), and the consideration of a finite tag-dictionary.

2.2. *Work in this context*

The challenge addressed in this paper concerns the quality and relevance of recommendations. The studies presented previously, based on a global evaluation of products, show some limits. Even if sometimes objective data like age or geographical data can improve the coherence of responses, it does not consider the perceptive reasons why the product is liked. We want to demonstrate the contribution of emotions in product evaluation. As our application case is a cultural product, film ; emotions are even more influential. Several studies use this application case such as (Saia, Boratto, and Carta 2014, 2016). In this work, the main objective is to eliminate the noise that may exist in the information of a

user account (history of a "sporadic" user, past trends) to improve the quality of recommendation. Based on semantic analysis of words in their context, it removes items that do not reflect the user's preference. The difference with our work is the passive source of information used (the user's history), compared to our expressed preference by the user. For this study, we consider that the information about a user concerns only the user himself. Three hypotheses support the proposal:

- H1: preference, represented here by the overall rating, is explained by customer emotions.
In (Poirson, Da Cunha, and Petiot 2014), we identify a link between emotions and preferences, to validate H1. In this paper, we determined emotions, of equal relevance, that explain the global rating of products. We decided to use the same list here, adapted to our product. To collect the emotions simply, we also chose to use the evaluation scale.
- H2: considering emotions improves recommendation quality.
The study of preference evaluation can give recommendation to a customer. (H1) implies that preference is influenced by emotions. The hypothesis we want to confirm is that a cluster of customers around a particular person, built on emotional data will contain closer neighbors in terms of preference.
- H3: customers do not have the same "preference profile".
'Emotions explain preference' does not mean that we have all the same profile of evaluation. We consider the existence of preference profiles: people with the same emotional evaluation can have the same preference (same profile) or not (different profile). In contrast, different perceptions can lead to similar preferences (for example, a person may like a film for its actors while another likes the same film because of its genre). We thus consider that clustering based on emotions and preference together can be less efficient than when based only on emotions.

The product chosen to illustrate our approach is film, a product for which preferences are mostly guided by emotions.

3. Recommendation methods

3.1. *Related work*

Recommendation systems rely on databases filled in by users. Whether it is spontaneous or provoked, the customer can give an assessment of his/her satisfaction with the services he/she has consumed. In order to express if she/he is satisfied or not, a global evaluation is expected. This global score is the result of an individual judgment of the interaction with the product, based on implicit data specific to each.

To complete this evaluation, our main hypothesis is as follows: "Taking into account customer's emotions improves recommendation quality". What is seen, heard, tasted and felt determines the product emotional impact (Fenech and Borg 2007). In order to transmit what is perceived, we chose to use evaluation based on semantic adjectives. The adjectives are taken from a list of basic emotions: joy, surprise, anger, disgust, contempt, shame, guilt, fear, interest, and sadness (Izard 1977), and adapted to the application case. Figure 1 illustrates the matrix of individual evaluations: the rows correspond to the different existing products, tested/evaluated or not by the customer, and the columns show the preference and emotions on which the product is evaluated. The number of customers gives the third dimension of the database.

The use case considered is: 'the system spontaneously recommends products to a customer called 'an active customer'.

The difficulty of the "sparsity problem" (Claypool et al. 1999) or the "gray sheep" (Sarwar et al. 2000) problem must be underlined. The sparsity problem occurs when there are not

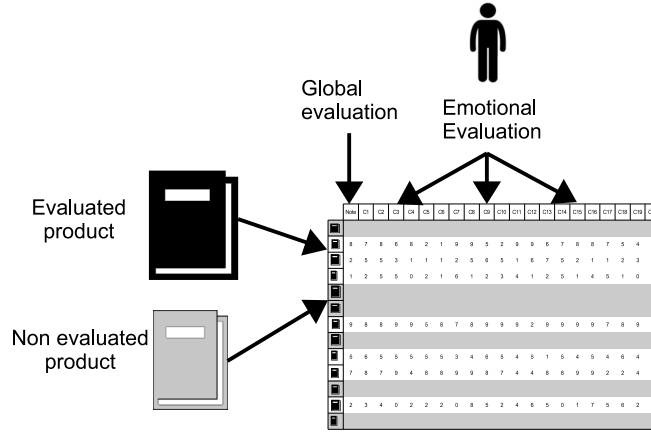


Figure 1. Format of an individual matrix of evaluation

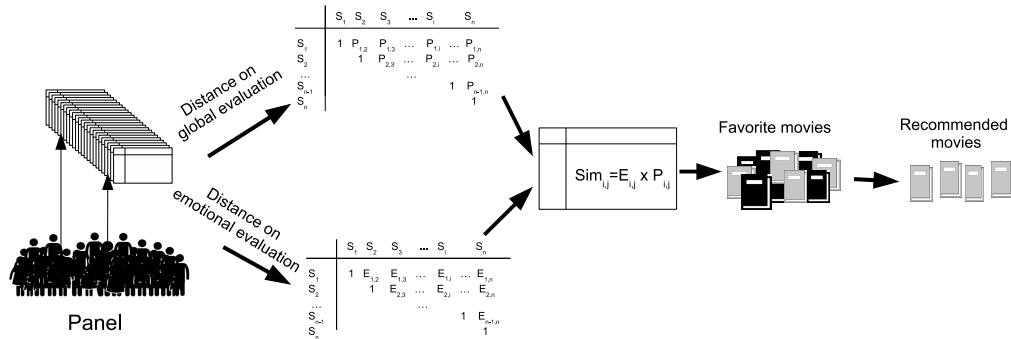


Figure 2. Presentation of the proposed approach

enough common products rated by customers. For example, if the number of customers is small relative to a large number of products, the matrix containing the customers product ratings could be sparse. It could be hard to find correlations between customers and consequently to find neighborhoods. The "gray sheep" problem occurs in a small community of customers, when there are customers with no close neighbors and whose ratings do not match any groups. The study will give some indications on the significance threshold of recommendation, depending on the matrix filling rate.

3.2. Proposed approach

The proposed approach is described in Figure 2. To recommend an item to a given subject:

- (1) Similarities are calculated between the considered subject and all the others,
- (2) A neighborhood (cluster) is built, encompassing all subjects within a similarity threshold,
- (3) For all unrated items, predictions of overall ratings are generated, and the best products are recommended,
- (4) The performance of the recommendations is assessed.

The following section describes our choices for the recommendation metrics.

3.2.1. Proposed similarity

In the traditional approach, a vector of preference evaluation represents a user. In our proposal, the database is composed of n sparse matrix (costumers) of m products evaluated (or not) on a global point of view (p) and on kem emotional attributes (evaluation $e_1 \dots e_{kem}$). Simultaneously, two similarities between the active user and the others are calculated: depending on the global evaluation and on the emotional scores.

The statistical tool used is the Pearson coefficient, which answers the question of the existence of a linear relationship between two sets of data. Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviations (Equation 1 for 2 variables X and Y of n values).

$$r_{xy} = \frac{cov(X, Y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^m (y_i - \bar{y})^2}} \quad (1)$$

with X and Y the 2 variables (users), m the number of products.

The result is between -1 (perfect negative correlation) to 1 (perfect positive correlation), 0 representing two independent variables.

The number of products must be discussed. Intuition can lead us to think that the greater the number of products in common, the more stable the prediction. Therefore, even if users have few products in common, but with identical evaluations, there is no reason to degrade their similarity. In statistical terms, the size of the sample is large. The more data we have, the more accurate the model is and the closer it gets to reality. In our case study, the data processed is perceptive or hedonistic. One cannot consider that there is "reality". The quantity of products tested is, therefore, not necessarily a decisive argument. Moreover, the proposed system operates on a dynamic database, with each new evaluation adding value. The effect of the "experiment" is therefore taken into account in Pearson's similarity assessments. We decided that the number of products in common will not influence the results of the recommendation.

However, let us consider the case of aberrant values, that may be due to a simple error during the evaluation for example.

Pearson's correlation coefficient is sensitive to extreme or aberrant points. In order to not over-penalize too much possible typing errors due to lack of attention, we, therefore, decided to ignore products whose evaluation difference is far from the mean, considering the threshold as the triple of the standard deviation (Eq.2).

$$\delta_{xy}^i = \left| |P_{x,i} - P_{y,i}| - |P_x - P_y| \right| - 3\sigma_{|P_x - P_y|} \quad (2)$$

with δ_{xy}^i the acceptable distance between x and y on the i^{eme} product, σ the standard deviation. For $\delta_{xy}^i > 0$, the product i is ignored. This first part gives a matrix of similarity in terms of preferences, $P_{X,Y}$.

The same operation is performed on the matrix of emotional evaluation: each user is represented by a matrix $E_{X,Y}$ similarity on emotional scores, one product tested by line. A measurement of the proximity between two customers, considering both preference and emotion is given, as the result of the multiplication of the first indicator ($P_{i,j}$, matrix of correlation on the preference evaluation) and the second one ($E_{i,j}$, matrix of correlation on the emotional evaluation) (Eq.3). The similarity is represented by $E_{i,j}$ when emotion rating similarity is considered, by $P_{i,j}$ when preference rating similarity is considered, and by their products when both are considered.

$$\begin{aligned}
\text{Preference similarity: } Sim_{X,Y} &= P_{X,Y} \\
\text{Emotion similarity: } Sim_{X,Y} &= E_{X,Y} \\
\text{Both similarity: } Sim_{X,Y} &= P_{X,Y} * E_{X,Y}
\end{aligned} \tag{3}$$

3.2.2. Proposed prediction

To recommend a product o_1 to an actual user S_i , an estimation of the preference is given (Eq.4). This estimation relies on the preferences of the other subjects, weighted by their similarities to the active customer. The mean preference is set as a basic evaluation in order to integrate the potential off-set in the scale used.

$$P_{S_i,o_1} = \left[\bar{P}_{S_i,o(o \neq o_1)} + \frac{\sum_{j=1, j \neq i}^n Sim_{S_i,S_j} * (P_{S_j,o_1} - \bar{P}_{S_i,o(o \neq o_1)})}{\sum_{j=1, j \neq i}^n Sim_{S_i,S_j}} \right] \tag{4}$$

with S_i the actual user, o_1 the studied product, n the number of users and Sim the matrix of similarity.

3.2.3. Proposed performance indicators

In order to assess the recommendation quality for customers, we defined an index Q . This index represents the global performance of the algorithm and is computed for each step of the cross-validation. The principle of one test is as follows:

- (1) Select randomly one of the considered subjects. This subject is the active one. Discard randomly an evaluated product,
- (2) Apply our approach to the remaining data,
- (3) Compare the real value with the estimation.

This procedure is repeated for all the subjects and all the evaluated products, one by one (660 * 100 tests). A removed product is considered "good", "bad" or "satisfactory" according to the overall rating of the active customer. A product is considered as

- "good" if the overall rating is in the upper quartile of this customer's evaluations,
- "bad" if the overall rating is in the lower quartile of this customer's evaluations.

We call:

G : the number of products good-rated removed from the system

B : the number of products bad-rated removed from the system

$C1$: number of good products with a good calculated evaluation ($C1 \leq G$)

$C2$: number of bad products with a good calculated evaluation ($C2 \leq B$)

The following expressions are proposed to compute the index Q (Formula 5), $Q \in [-1, 1]$

$$Q = \begin{cases} 0 & , \text{ if } B = G = 0, \\ -\frac{C2}{B} & , \text{ if } G = 0 \text{ and } B \neq 0, \\ \frac{C1}{G} & , \text{ if } B = 0 \text{ and } G \neq 0, \\ \frac{C1}{G} - \frac{C2}{B} & , \text{ otherwise.} \end{cases} \tag{5}$$

the mean distance between real preference and predicted can also be calculate.

4. Experimentation

4.1. Experimental protocol

Our experiment comprised two steps: a first step using simulated data and a second test using real data.

The objective of the first step was to test in a controlled environment (where a link exists between emotions and preference) that using emotions improves the recommendation (assumption H2). This step also aimed to verify the impact of different parameters on the performance of the proposed approach. In particular, the impact of sparsity and noise of the input data on performance should be stressed.

The objective of the second step was to verify our assumption that using perceptions instead of preference improve recommendation for real subjects.

4.2. Simulated case description

In order to test our assumptions in a controlled environment, data is simulated. We created a pool of 660 customers and 100 items (thus 66 000 tests). To be coherent with human behavior, we decided not to generate the data completely randomly: for a family of products, it is rare to judge them in completely different ways. We decided empirically to create 3 emotion profiles and 4 preference profiles thus 12 different subject profiles.

4.2.1. Emotion profiles

We defined 3 different emotion profiles, i.e. evaluation for 8 emotions per item, noted $e_{i,o,k}$ with i the customer, o the item evaluated and k the emotion.

- Profile A: the 8 emotion evaluations are randomly attributed (uniform law $[0,10]$).
- Profile B: the customer not use all the scale, the 8 emotion evaluations are randomly attributed (uniform law $[3,7]$).
- Profile C: the customer evaluates using the extreme graduations (Beta law, $\alpha = \beta = 0.5$)

Figure 3 presents the probability density and the cumulative distribution functions. Reading figure 3(b), we see that the probability for an evaluation lower or equal to 7 is 70% for profile A, 100% for profile B and 63% for profile C.

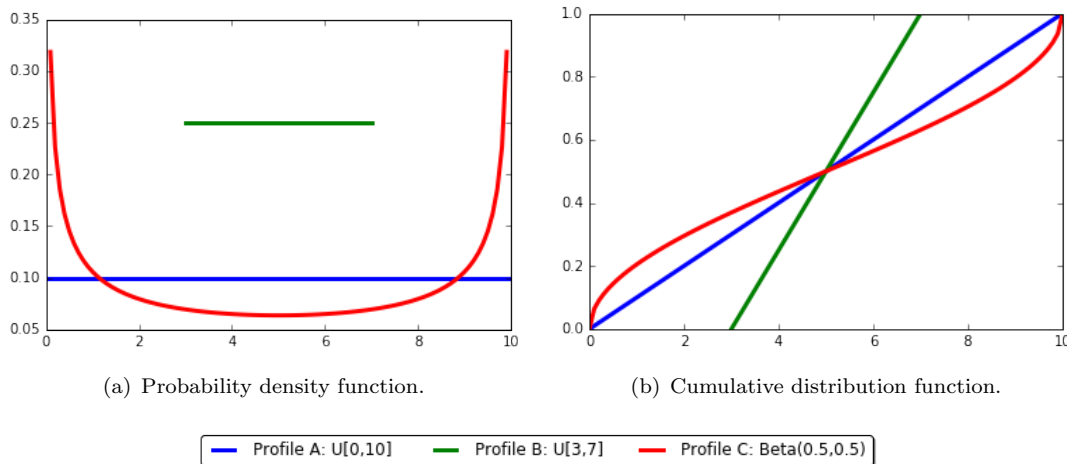


Figure 3. Probability distributions used for data generation

	Pref.	emotion 1	emo. 2	emo. 3	emo. 4	emo. 5	emo. 6	emo. 7	emo. 8
<i>Bob's emotional profile</i>		1	1	1	1	1	1	1	1
Bob's evaluation	5	4	5	6	4	7	6	5	3
<i>Max's emotional profile</i>		2	3	6	5	8	6	6	2
Max's evaluations	5	2	8	3	9	3	9	3	2

Table 1. Example of ratings of the same item for 2 customers: same preference for different emotions

4.2.2. Preference profiles

We defined 4 different preference profiles. They consist in 8 coefficients (noted $w_{i,k}$ with i the customer and k the emotion), weighting the emotional evaluations to obtain the overall evaluation: the preference. For the 4 profiles, the coefficients were randomly chosen using a uniform law ($U[1,10]$). A Gaussian noise term is added to mimic human behavior. In order to keep a leveled scale, the integer part is used.

$$P_{s_i,o} = \left\lfloor \frac{\sum_k w_{i,k} \times (e_{i,o,k} + noise \times N'(0,1))}{\sum_k e_{i,o,k}} + noise \times N(0,1) \right\rfloor \quad (6)$$

Figure 4 represents the obtained preferences.

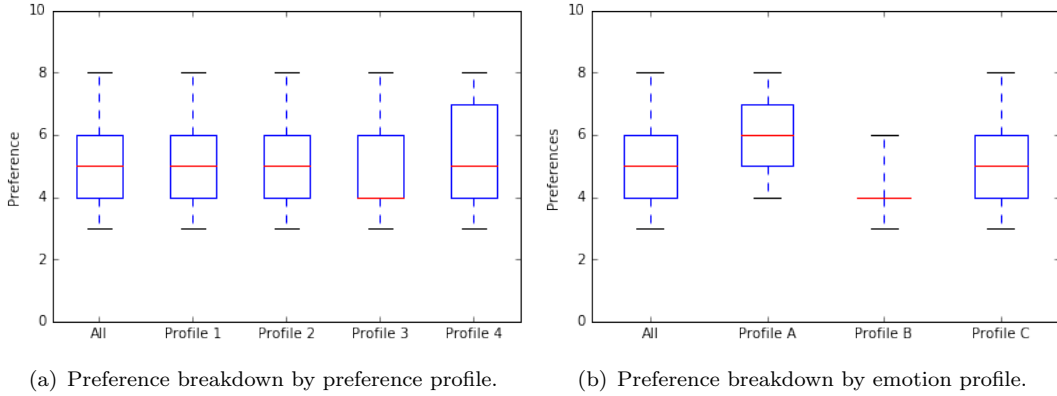


Figure 4. Preference breakdown by profile.

Figure 4(a) stresses the fact that for different emotional evaluations, the preference can be the same. As numerous combinations of emotions can give the same global rating, differentiating between preference profiles is difficult (cf. example in table 1). On the contrary, figure 4(b) illustrates that emotion profiles have a huge importance on the preference evaluation, regardless of the preference profile. People are more distinguishable in their emotions than in their preferences.

4.3. Real case description

4.3.1. Application sector

The product chosen to illustrate our approach is film, a product for which preferences are mostly guided by emotions. The first step is to determine a list of emotional attributes, adapted to the product. These attributes were identified from information given by the Differential Emotions Scale (DES) (Izard 1977), and adaptations of the DES (Philippot 1993; Gross and Levenson 1995). The list is composed of: Amused, Sad, Angry, Scared, Attentive, Anxious, Surprise and Happy.

For each film seen, the customer provides two types of quantitative evaluation:

- An overall rating of the film, that corresponds to the overall customer preference for the product. This rating is performed on a numerical scale of 10 levels.
- Rating of the film based on emotional criteria, using a quantitative scale of 7 levels (McHugo, Smith, and Lanzetta 1982).

4.3.2. The online questionnaire

To define the list of the films presented in the questionnaire, nine film fans were asked to list exhaustively the films they have seen in the last three years. From this selection, a final list of 100 films was defined. The films were selected so as to cover a wide range of genres (adventure, comedy, fantasy, animation, drama, action, horror ...). In order to collect ratings data, the questionnaire was published online. Respondents had the choice to leave the questionnaire at any time (if they felt bored). They also indicated their gender and age. The target population was expected to have a good understanding of French (the language of evaluation). If the respondent indicated that they had seen a film, he/she was redirected to the page corresponding to its evaluation.

4.3.3. Information on respondents

562 respondents were considered relevant to our study (more than 4 films seen). For these respondents, the average number of films rated was 13.07. Among them, we distinguished 186 women and 376 men; 501 respondents reported their ages, including 451 under 30 years of age and 50 respondents over 30 years of age.

The data collected was formatted in a 3D table: subject (562)*films(100)*global and emotional evaluations (1+8). The matrix had a sparse density: the number of respondents by film was variable, which is a good representation of reality. For example, the 25th film (Intouchables by Nakache&Toledano, a French blockbuster from 2011) had been seen by almost all subjects, whereas some had been seen by only twenty people. The data was more scattered towards the end of the list. This shows a potential limit of the test, the length of the questionnaire being perhaps superior to the patience of the testers. The order of presentation of the films should have been more varied.

4.4. Computational environment

Computational experimentation was performed using Python 3.1 programming. The computational expense of the experiment (time and memory) was not evaluated, since for the real application the calculation of the recommender system would be done offline.

5. Results and discussion

As a reminder, for each of the 660 customers, all preference evaluations, were deleted one by one and simulated based on the preferences of the other customers, or their emotional evaluations. The result was compared to the original value.

5.1. Simulated case results

Our simulated data complied with our hypothesis of a link between emotional evaluation and preference (H1). We used this simulated data to assess the performance of our recommendation method and the relevance of our performance indicator, and, therefore, to validate the hypothesis H2.

The first point to evaluate is the influence of matrix sparsity. We, therefore, tested the performance of the recommendations, randomly removing evaluations. The number of removed evaluations per customer informs the indicator of sparsity, ranging from 0 (all customers evaluated all products) to 80 (all customers evaluated only 20 products out of 100). Figure 5(a) represents the distance between the two evaluations (the discarded and the calculated) on 5 recommendation methods: on emotional similarity, on preference similarity, on both similarities ($P * E$), on the preference mean on the other products tested (considering that a user evaluates all the films identically) and a randomly generated evaluation. Firstly, the difference between random and other methods shows the importance of studying user evaluations in detail. Secondly, we noticed that the recommendations based on emotions gave better results than the other recommendations. We were also able to deduce that sparsity does not have a significant influence on the efficiency of the recommender. A high sparsity ($\geq 70\%$) moderates the preference rating performance because the less data, the more important the standard deviation is, implying a probable error. This is less the case for the data of emotions because of its vector nature. The distance indicator highlights the fact that using the mean value of preferences to predict preference allows the error to be managed but does not allow the actual appeal for a given product to be predicted (translated here by a preference in the upper quartile). That is why the Q value obtained by this method is around zero (cf. figure 5(b)). Figure 5(b) represents Q, the performance indicator of our recommendation approach. Even when the matrix had high sparsity, our recommendation was efficient. We decided to challenge the recommender at a limit value of 80% and use this sparsity for the following tests.

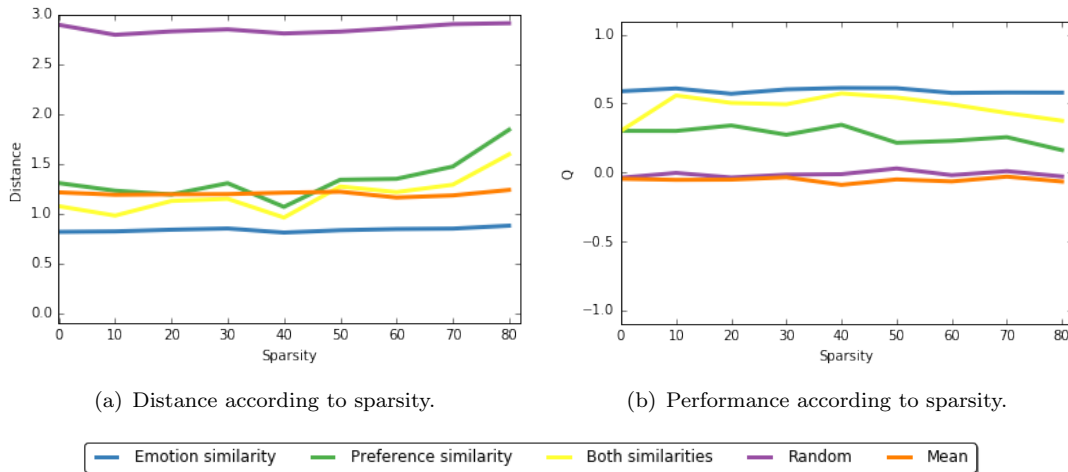
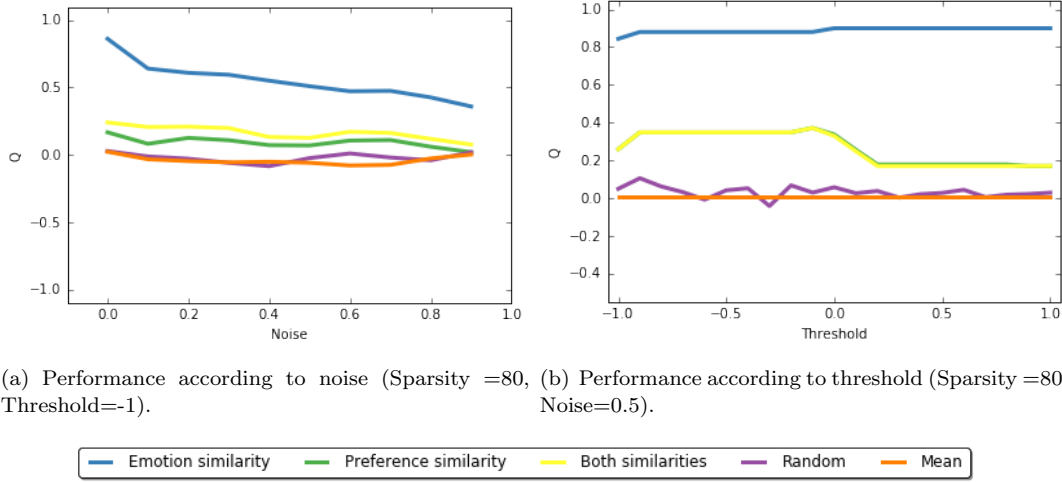


Figure 5. Impact of sparsity (Noise =0.5, Threshold=-1).

The second analysis was on the influence of noise added to mimic human behavior. The noise impact on our preferences is stated in equation 6. Logically, the greater the noise, the closer the estimations to a random projection. Large noise erased our user profiles. The results presented in figure 6(a) show that the quality of the recommendation reduces when noise increases. The performance of the emotion similarity recommendation remained better than the other methods. However, to keep the diversity of the population, we chose a noise equivalent to 0.5 for the other tests.

In a traditional recommendation system, only "similar" customers are considered to predict the behavior of the considered one. In our case, the similarity was evaluated through the Pearson correlation (cf. section 3.2.1). We tested the impact of limiting the information used in the recommendation to only similar customers, applying a threshold to the similarity. The results presented in figure 6(b) show that recommendations are better when all information is considered. If someone who displays a behavior opposite to your own loves a film, you

could infer that you will not love it at all. We chose to keep all customers, setting the threshold to its smallest value: -1.



(a) Performance according to noise (Sparsity =80, Threshold=-1). (b) Performance according to threshold (Sparsity =80, Noise=0.5).

Figure 6. Impact of noise and threshold.

The results of the simulated case can be summarized as follows:

- emotion similarity gives better recommendations than preference similarity, irrespective of sparsity and noise,
- sparsity has only a small impact on the performance of recommendations based on emotion,
- noise impacts the performance of recommendations based on emotion,
- recommendations based on emotion are more efficient when considering all subjects.

5.2. Real case

For the real data the test procedure was an **all-but-one** performed on all subjects and all films. Due to sparsity, 7346 tests were done.

The similarity threshold chosen was -1 (all the other subjects are considered).

The recommendation using emotion similarity obtained the best performance both for distance and quality indicators (bold numbers in table 2). Using the mean preference as a predictor of future preferences turned out well on a distance level (performance of 1.3, better than that of preference similarity, 1.42) but it proved unable to anticipate the loved (upper quartile) and the hated (lower quartile) films - the quality indicator being 0.01.

The results obtained with emotional similarity were better than those using both preference and emotions; this stresses the fact that adding input information does not suffice to improve the recommendation.

	Emotion similarity	Preference similarity	Both similarities	Random	Mean
Quality	0.25	0.12	0.12	0.03	0.01
Distance (real - calculated)	1.06	1.42	1.40	2.71	1.30

Table 2. Performance of the different methods

Building on the results of the simulated case, the results of the real case can be summarized as follows:

- adding information does not suffice to improve the recommendation,
- emotion similarity gives better recommendations than preference similarity, for quality and distance

This reinforces the existence of 'preference profiles' that can be effectively used for recommendation (H3).

6. Conclusion

This paper discussed the use of emotions to improve the quality of product recommendation. The initial hypothesis was that emotions are fundamentals in the act of purchase, especially for products related to the cultural field. We proposed a collaborative method. The opinions of users of the same community allow us to predict product evaluations for a given user. He/she will thus be advised of the relevant product to purchase. To test our proposal, we generated a database of customers/products/evaluation (preference and emotions). A previously removed data item was simulated, based on other customer evaluations, and compared to the original value. The results obtained on the simulated data validate our hypothesis that if emotions influence preference, then using emotional similarity improves the product recommendations. The second important point is that the proposed method is not very sensitive to database sparsity. That is important because in reality, emotion databases are rarely exhaustively complete. The second step of the study was a real case study: an online questionnaire on films. The results obtained with our method on real data were better than the one using only preferences. This validated our hypothesis that the preference expressed by customers for a given product is the result of product/customer interaction, which is translated into emotions. We believe that consideration of emotions should thus be factored into all recommendation systems (collaborative filtering, hybrid filtering, etc.). In further research, a methodology systematically obtain emotional descriptors for any type of product could be studied. The challenge is to improve emotion collection. In our study, the user was asked to evaluate his/her feelings on scales labeled with an emotion, which is not an easy task. The solution envisaged is to work on the semantics of user reviews, extracting from these comments the emotional data which enters into the purchase decision and to compute them with our recommender proposal. Another direction is also to use facial emotion detection to simply (for the customer) complete a matrix in real time.

7. Acknowledgement

The authors are grateful for the work of Ines Jomaa, ex-PhD student, in the administration of the online questionnaire.

References

- Al-Shamri, M.Y.H. 2016. "User profiling approaches for demographic recommender systems." *Knowledge-Based Systems* 100: 175–187.
- Claypool, M., A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin. 1999. *Combining Content-based and collaborative filters in an online newspaper*. Technical Report. Worcester Polytechnic Institute.
- Degemmis, M., P. Lops, and G. Semeraro. 2007. "A content-collaborative recommender that exploits WordNet-based user profiles for neighborhood formation." *User Model User-Adap-Inter* 17: 217–255.
- Deng, Shuiguang, Dongjing Wang, Xitong Li, and Guandong Xu. 2015. "Exploring user

- emotion in microblogs for music recommendation.” *Expert Systems with Applications* 42: 9284–9293.
- Desmet, P.M.A., R. Porcelijn, and M.B. van Dijk. 2007. “Emotional Design; Application of a Research-Based Design Approach.” *Know Techn Pol* 20: 141–155.
- Englern, Tobias H., Patrick Winter, and Michael Schulz. 2015. “Understanding online product ratings: A customer satisfaction model.” *Journal of Retailing and Consumer Services* 27: 113–120.
- Fenech, O.C., and J.C. Borg. 2007. “Exploiting emotions for successful product design.” In *Proceedings of International Conference of Engineering Design ICED’07*, .
- Gross, J.J., and R.W. Levenson. 1995. “Emotion Elicitation Using Films.” *Cognition and Emotion* 9 (1): 87–108.
- Herlocker, J., J.Konstan, and J.Riedl. 2000. “Explaining collaborative filtering recommendations.” In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, 241–250.
- Izard, C.E. 1977. *Human Emotions*. New York: Plenum Press.
- Jannach, D., and G. Kreutler. 2007. “Rapid development of knowledge-based conversational recommender applications with advisor suite.” *Journal of Web Engineering* 6 (3): 165–192.
- Jannach, D., M. Zanker, and M. Fuchs. 2014. “Leveraging multi-criteria customer feedback for satisfaction analysis and improved recommendations.” *J. Inf. Technol* 14 (2): 119–149.
- Konstan, J.A., and J.Riedl. 2012. “Recommender Systems: From Algorithms to User Experience.” *User Modeling and User-Adapted Interaction* 22: 101–123.
- Li, Lei, and Tao Li. 2013. “News Recommendation via Hypergraph Learning: Encapsulation of User Behavior and News Content.” In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, WSDM ’13*, New York, NY, USA, 305–314. ACM. <http://doi.acm.org/10.1145/2433396.2433436>.
- Li, Lei, Li Zheng, Fan Yang, and Tao Lia. 2014. “Modeling and broadening temporal user interest in personalized news recommendation.” *Expert Systems with Applications* 41 (7): 3168–3177.
- Lim, Hyewon, and Hyoung-Joo Kim. 2017. “Item recommendation using tag emotion in social cataloging services.” *Expert Systems with Applications* 89: 179–187.
- McHugo, G.J., C.A. Smith, and J.T. Lanzetta. 1982. “The Structure of Self-Report of Emotional Responses to Film Segments.” *Motivation and Emotion* 6 (4): 365–385.
- Philippot, P. 1993. “Inducing and Assessing Differentiated Emotion-Feeling States in the Laboratory.” *Cognition and Emotion* 7 (2): 171–193.
- Poirson, E., C. Da Cunha, and J. F. Petiot. 2014. “Can we categorize moviegoers on their emotions?” In *Kanseï engineering and emotion research-KEER*, 1216–1219.
- Porcel, C., A. Tejada-Lorente, M. Martínez, and E. Herrera-Viedma. 2012. “A hybrid recommender system for the selective dissemination of research resources in a technology transfer office.” *Inf. Sci.* 184 (1): 1–19.
- Richins, M. 1997. “Measuring emotions in the consumption experience.” *Journal of Consumer Research* 24 (2): 127–146.
- Saia, Roberto, Ludovico Boratto, and Salvatore Carta. 2014. “Semantic Coherence-based User Profile Modeling in the Recommender Systems Context.” *KDIR* .
- Saia, Roberto, Ludovico Boratto, and Salvatore Carta. 2016. “A semantic approach to remove incoherent items from a user profile and improve the accuracy of a recommender system.” *Journal of Intelligent Information Systems* 47 (1): 111–134. <https://doi.org/10.1007/s10844-016-0406-7>.
- Sarwar, B., G. Karypis, J. Konstan, and J. Riedl. 2000. “Analysis of Recommendation Algorithms for E-Commerce.” In *2nd ACM Conference on Electronic Commerce*, 158–167.
- Serrano-Guerrero, J., E. Herrera-Viedma, J.A. Olivas, A. Cerezo, and F.P. Romero. 2011. “A google wave-based fuzzy recommender system to disseminate information in University Digital Libraries 2.0.” In *Inf. Sci.*, Vol. 181(9), 1503–1516.
- Shang, H., C. Hsu, Ming, C. Chuang, Chien, and C. Chang. 2000. “A semantic differential study of designers’ and users’ product form perception.” *International Journal of*

- Industrial Ergonomics*. 25 (375-391).
- Tkalčič, M., A. Kosir, and J. Tasic. 2011. "Affective recommender systems: The role of emotions in recommender systems." *Proceedings of the RecSys 2011 workshop on human decision making in recommender systems, Citeseer* 9–13.
- Winoto, Pinata, and Tiffany Y. Tang. 2010. "The role of user mood in movie recommendations." *Expert Systems with Applications* 37 (8): 6086 – 6092.
<http://www.sciencedirect.com/science/article/pii/S0957417410001569>.