Emotive and Personality Parameters in Multimedia Recommender Systems

Marko Tkalčič, Jurij Tasič and Andrej Košir
University of Ljubljana Faculty of Electrical Engineering
Tržaška 25, Ljubljana, Slovenia
{marko.tkalcic|jurij.tasic|andrej.kosir}@fe.uni-lj.si

Abstract

The PhD research presented in this paper discusses the inclusion of affective computing in multimedia recommender systems. Our objective is to fill the missing gaps of the scenario where the system detects the emotive response of a user watching a multimedia item. The emotive state is then used to model items and users in the recommender system which filters only relevant items for each specific user. The thesis is composed of four original scientific contributions: (i) a module for emotion detection from video sequences of user faces into the valence-arousal-dominance (VAD) space, (ii) a content-based recommender system with affective user and item modelling, (iii) a collaborative filtering recommender algorithm with a personality-based user similarity measure and (iv) a database that forms the basis for the first three contributions. The paper gives results for contribution (ii) and preliminary results for contribution (iii). Contribution (iv) is also presented while a tentative future plan for contribution (i) is given.

Keywords: content-based recommender system, collaborative recommender system, emotive response, emotion detection, big five personality model

1 Introduction

The presented PhD research addresses affective computing in multimedia recommender systems (MRS). Recommender systems are getting more and more attention as the amount of available multimedia content is growing. The goal of MRS is to make a narrow selection of multimedia content that is relevant for each specific user based on her/his preferences. The user is thus more satisfied with the system. Although this is a task that should intuitively follow the end user's affective responses and personality, state of the art MRS mostly ignore the affective approach. Most of today's MRS create the user profiles (a data structure that contains the knowledge about a user) based on past user's actions. This user feedback can be explicit (e.g. ratings) or implicit (observing the user's behaviour in a non-intrusive manner). In laboratory experiments we usually rely on explicit feedback which is more accurate but in real life applications it is undesired because it is too intrusive and tends to turn users away. Thus an implicit feedback collection technique is preferable.

1.1 Problem Statement and Proposed Solution

Except for few research contributions (González et al., 2004; Lekakos and Giaglis, 2006; Shan et al., 2009), MRS do not exploit the affective aspect in human computer interaction (HCI). The two most popular MRS approaches, content-based recommenders (CBR) and collaborative filtering recommenders (CF), both use very technical information to perform the prediction of how a user would rate (and like) an item.

We believe that bringing emotions and personality in the field of user modelling for multimedia recommender systems would improve their performance. We propose a user scenario (inspired by video-on-demand like services) where end users consume multimedia content, their affective responses are detected in a non intrusive fashion and used to build the user and item profiles for the CBR system. As an alternative to the CBR we propose to have a CF recommender where the user similarity measure is calculated based on end users' personalities. The user is then offered a narrow selection of relevant items instead of browsing the whole database. Figure 1 shows the proposed scenario. We want to get closer to the system predicted by Picard (2000): *...that when the machine presents you with something you like, it sees that you like it. And when it does something you do not like, it sees that too [page 101].*

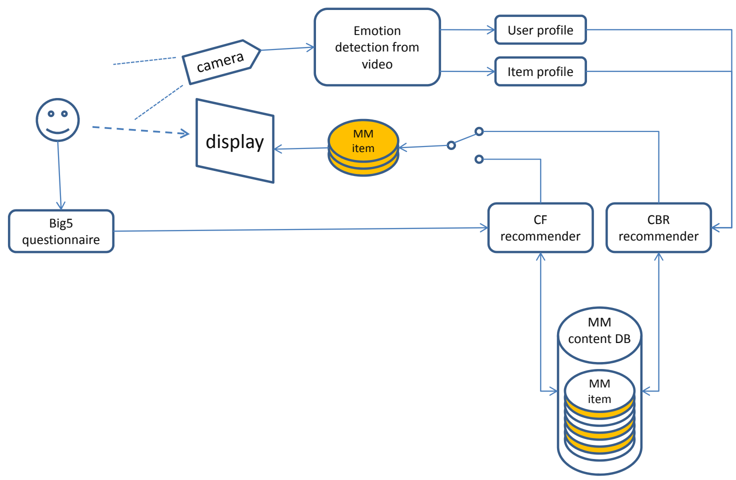

Figure 1: The recommender system scenario: while the user is consuming an item (movie, picture etc.) the system detects her/his emotive response in a non-intrusive way. This information is used to build the profiles for the CBR system which retrieves only the items relevant for that user from the database. As an alternative to the CBR, a CF recommender that exploits the users' personalities can be used to perform the filtering of relevant items.

1.2 Objectives and Scientific Contributions

Some elements of such a system already do exist but are being used for other purposes and need to be properly adapted while some others are missing. Our objective is to fill these gaps. Thus the original scientific contributions of this PhD research will be the following

Emotion detection from video A non intrusive emotion detection technique that maps video sequences of users' faces into emotive states in the valence-arousal-dominance (VAD) emotive space.

Affective item and user profiles for CBR systems A CBR that uses information about the user's emotive response during the content consumption as metadata for modelling both items and users (instead of using genre based metadata).

Personality based similarity measure for CF systems A CF recommender that calculates the user similarity using a personality based measure (instead of using rating based measures).

Dataset of user interaction with affective metadata and explicit ratings A database of user interaction with a multimedia consumption system that contains explicit ratings, affective responses in the VAD space, users' big five personality values and video sequences of the users while consuming the items.

By fulfilling our goals we wish to bring the affective computing approach in the area of multimedia recommender systems where we believe it should be present because the entertainment business is primarily about consumer's emotions. This research work addresses several standards, in particular the W3C EmotionML (see Schröder et al. (2008)) and MPEG7 Usage History (see Manjunath et al. (2002)).

1.3 Related Work

A lot of work has been done in the field of detecting emotions from various modalities, see Donato et al. (1999); Chibelushi and Bourel (2003); Fasel and Luettin (2003); Picard and Daily (2005); Zeng et al. (2009) for excellent overviews. All of these methods yield one of the basic emotions as output.

Overviews of recommender systems are given by Adomavicius and Tuzhilin (2005); Burke (2002); Lew et al. (2006). The basic taxonomy is given (content-based, collaborative and hybrid) and the pros and cons are described. An example of a CBR system is the work done by Pogačnik et al. (2005) while Kunaver et al. (2007) have been working on CF and hybrid methods.

There have been attempts at introducing affective modelling of end users in recommender systems. Lekakos and Giaglis (2006) took the user's lifestyle as the central parameter of their collaborative recommender. González et al. (2004) introduced the Smart User Model (SUM) which is a data structure composed of objective attributes, subjective attributes and psychological traits.

Several authors tried to annotate multimedia content with information about the emotive state an item induces in end users. Hanjalic (2006) extracted mood from video sequences. Lang et al. (2005) performed a large scale experiment with a big image database and several subjects who manually annotated each image with their emotive response which yielded the IAPS database.

There are several taxonomies for the description of emotions (Cowie et al., 2001; Posner et al., 2005; Villon and Lisetti, 2006). One line follows the work started by Charles Darwin and is called the *basic emotions theory* (Ekman and Friesen, 2003). The other popular description of emotions in HCI is the three dimensional space valence - arousal - dominance (VAD), where each emotive state is represented by a triple of numerical values that describe a certain qualitative aspect of the emotion. The *circumplex model of emotions* maps the basic emotions into the VAD space (Posner et al., 2005; Villon and Lisetti, 2006).

2 Methodology

In this section we present the methodology for each of the four scientific contributions. We provide the hypotheses, describe the plan of work and/or work already done, give results where available and provide a list of open issues for each contribution.

The basis of all the contributions is the database of users consuming digital items (images) which is described in more details in Sec. 2.4. We collected the big five personality values for all users, tracked their explicit ratings for each image and recorded the emotive responses of their faces with a video camera. Each image was annotated with a genre attribute and a six-tuple of the induced emotive state.

2.1 Emotion Detection from Video Sequences

Our hypothesis is that it is possible to detect the emotive responses of end users from video sequences of their faces with a certain success rate. The target space of emotion description is the VAD space.

At the time of writing we have not started yet with the experimental part of the emotion detection from video. However we performed a preliminary literature review (Wang and Guan, 2008; Kim and André, 2008; Zeng et al., 2009; Donato et al., 1999; Chibelushi and Bourel, 2003; Fasel and Luettin, 2003) which led us to the following methodological plan:

1. Face detection in all video sequences using the Viola-Jones algorithm (Viola and Jones, 2004) implemented in OpenCV

2. Registration of faces to improve the Viola-Jones detection results. We will start with the template matching technique (Matlab)
3. Feature extraction: based on literature review we will first try with Gabor filtering on each frame and use the Hidden Markov model HMM to produce length invariant features (Matlab)
4. Merge the features of video sequences and emotion labels (VAD values) into the dataset
5. Evaluate various machine learning algorithms with the dataset. We will use the 10 fold cross validation scheme to build the training and test sets. (Matlab, Weka)
6. Perform statistical tests to assess the results yielded

2.1.1 Open Issues

As we have not started yet with emotion detection there are several open issues and many more that we are currently not aware of: (i) choice of features: which are the best features taking into account accuracy and speed of emotion detection, (ii) choice of the machine learning algorithm: which is the most suitable classifier, (iii) determination of the acceptable success rate and (iv) determination of the classifier's output: nominal or numerical classes.

2.2 Content Based Recommender with Affective User Modelling

The hypothesis of this contribution is that CBR algorithms with affective based metadata perform better than CBR with standard genre based metadata.

In order to accept (or reject) the above hypothesis we constructed two metadata sets for the item profiles: an affective and a standard (non affective) and compared the performance of a CBR using both metadata sets for item/user profiles.

Each multimedia content item induces an emotive response in end users. This response is described with the values v, a and d in the VAD space. The first two statistical moments of the responses of a set of users that have watched the same item h formed the affective metadata set $\mathcal{V}$ of the item h:

$$\mathcal{V} = (\bar{v}, \sigma_v, \bar{a}, \sigma_a, \bar{d}, \sigma_d) \tag{1}$$

which is a six-tuple. The standard metadata set A was composed of the genre and the average watching time for the item h. The genre was a value from a set of ten genres and the average watching time was calculated from different users for the observed item h.

2.2.1 Experiment and Results

We performed the prediction of item relevancy with machine learning algorithms. We evaluated the AdaBoost, NaiveBayes, C4.5 and Support Vector Machine (SVM) classifiers. We trained the classifiers with two metadata sets: A (affective) and $A \times \mathcal{V}$ (standard and affective) . We used the ten fold cross validation for splitting the dataset into the training and test sets. From the confusion matrices we calculated the precision P, recall R and F-measure as defined by Herlocker et al. (2004). We also performed the Pearson χ^2 significance test comparing the confusion matrices yielded by the metadata sets A and $A \times \mathcal{V}$.

The results of the CBR recommender with affective user modelling have been submitted for publication to a journal (Tkalčič et al., 2009). In table 1 we present the performance results of the different metadata sets and classifiers. The Pearson χ^2 significance test showed that the difference was significant in all four classifiers.

2.2.2 Open Issues

The results showed that modelling users based on their preferences for emotive responses gives better results than modelling with genre based parameters. But we are still unsure whether the approach taken could be improved by using the emotive response parameters in a different way.

Thus the open issues regarding this topic are: (i) is the *VAD* space better than the *basic emotions* space for MRS user modelling and (ii) in which way could the described approach be improved (multivariate analysis).

metadata set	classifier	P	R	F
A	AdaBoost	0,57	0,42	0,48
	C4.5	0,60	0,46	0,52
	NaiveBayes	0,58	0,58	0,58
	SVM	0,61	0,55	0,58
$A \times \mathcal{V}$	AdaBoost	0,63	0,56	0,59
	C4.5	0,64	0,57	0,60
	NaiveBayes	0,57	0,64	0,61
	SVM	0,65	0,61	0,63

Table 1: Precision, recall and F measure for the two metadata sets and four classifiers.

2.3 Collaborative Recommender with Personality-based User Similarity Measures

The hypothesis is that the usage of personality based user similarity measures yields statistically equivalent or significantly better results than the usage of rating based similarity measures in terms of the recommender system's performance (precision, recall, F-measure).

CF recommenders are based on the presumption that when the similarity between two users $sim(u_1, u_2)$ is high both users will give similar ratings to the item h. The user similarity measure is thus a crucial part of any CF system. The similarity measures used in state-of-the art CF systems are rating based. The main drawback is that it needs to be calculated on a regularly basis as new ratings are added to the system and the computational complexity is high.

We propose to use similarity measures whose calculations are fast and do not need to be recalculated on a regular basis. We evaluated the usage of a set of new similarity measures for a CF system based on the big five personality model, which describes the personality of a single user u by giving numerical values to five dimensions: openness, conscientiousness, extraversion, agreeableness, and neuroticism (Goldberg et al., 2006). We denote these with the vector $\vec{b} = (b_1, \ldots, b_5)$. The assessment of these values for a specific person is usually done through a questionnaire like the one proposed by Goldberg et al. (2006).

We evaluated two approaches for the calculation of distances between users u_1 and u_2 (which are inversely related to the similarity $sim(u_1, u_2)$) based on the users' personality vectors $\vec{b}_1$ and $\vec{b}_2$:

1. Euclidian distance: $d_E(\vec{b}_1, \vec{b}_2)^2 = \sum_i |b_{1i} - b_{2i}|^2$
2. Weighted Euclidian distance: $d_{WE}(\vec{b}_1, \vec{b}_2)^2 = \sum_i w_i^2 |b_{1i} - b_{2i}|^2$ where the weights w_i were the coefficients of the first vector yielded by the PCA

2.3.1 Experiment and Preliminary Results

We applied the acquired dataset (see Sec. 2.4) in the CF recommender system developed by Kunaver et al. (2007). Each similarity measure was evaluated two times: once with a higher weight on close neighbours and once with a higher weight on all users (public). The reference rating-based similarity measure (implemented by Kunaver et al. (2007); with which we compared the personality based measures) and the two big five measures thus yielded six different combinations of similarity measures to calculate and evaluate.

We used the precision P, recall R and F-measure as performance measures. We also performed a one-way analysis of variance to determine whether the differences of mean values F of all six measures were statistically significant.

In terms of mean values of P, R and F the big five based approaches performed better than the ratings based approaches. The mean values of P, R and F are reported in Tab. 2. The ANOVA analysis (with the significance level $\alpha = 0.05$) of the F measure further showed that all the big five based approaches performed significantly better than the distance based measure with

	P	R	F
distances - neighbours	0.6666	0.5895	0.6268
distances - public	0.7042	0.7401	0.7232
big5 neighbours	0.6309	0.8533	0.7062
big5 public	0.7093	0.8068	0.7442
weighted big5 neighbours	0.6455	0.8398	0.7165
weighted big5 public	0.7104	0.8064	0.7450

Table 2: Mean values of P, R and F for the different combinations of measures and weighting.

higher weight on the neighbours and were statistically equivalent to the distance based measure with higher weight on the public. The box plot of the F-measure values is shown in Fig. 2.

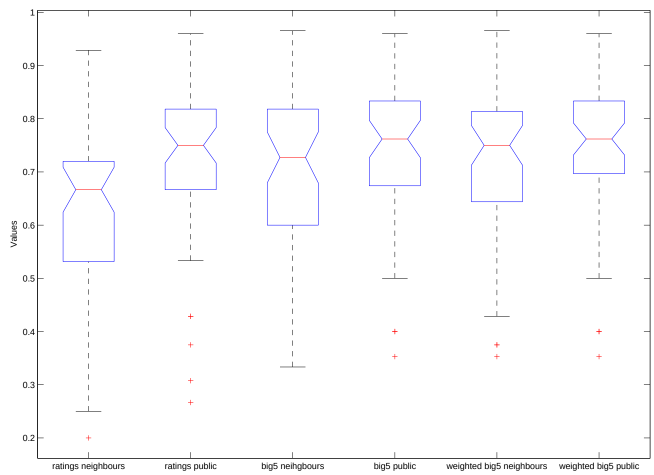

Figure 2: Box plots of the F measure results for six variants of measures.

2.3.2 Open Issues

The personality based measures seem to produce quality neighbours, which lead the CF recommender to results equivalent to the rating-based measures. However we identified three important issues pending: (i) in order to get the big five values end users must first fill in a questionnaire which is annoying. A better approach would be a solution similar to the one proposed by Berger et al. (2007) who found significant dependencies between photos and tourist types. Based on this they developed a game where end users drag a set of photos into a *like-it* bin and a *don't like-it* bin and the system classifies the user in a category of the kick-pack plane used in tourist profiling. We plan to investigate the relation between the ratings of specific images from our database and the personality values in order to develop a funny automatic personality detector. The second issue is (ii) the absence of context awareness (a user can have different neighbours in different situations, e.g. watching a movie with her/his friends vs. watching a movie with her/his kids) in the calculation of neighbours and the third is (iii) that we need to evaluate more similarity measures to give the results heavier foundations.

2.4 Database

Although the database has been the first methodological step that was performed we put it as the last subsection because it is easier to understand the requirements for the database at this point.

We needed a dataset in the form of a history log of interactions of users with a device displaying multimedia content. The following attributes were needed in the dataset: user's big five personality values, explicit ratings and induced emotive states.

We performed an emotion induction experiment. We chose 70 images from the IAPS database (Lang et al., 2005) with apriori known VAD values of the induced emotions. We had 52 users (15 males and 37 females) aged between 17 and 20 taking part in the experiment. Each user was shown a sequence of images and was requested to give an explicit rating from a five point Likert scale to each image. The users were recorded during the consumption of the images with a web camera. Furthermore each user filled in a questionnaire from the IPIP pool (http://ipip.ori.org/, 2009) in order to asses her/his big five personality values.

2.4.1 Open Issues

A further statistical analysis of the dataset is needed. In the future we plan to extend the database with: (i) more users, (ii) more items and (iii) contextual information. We plan to perform the experiment with each user repeating it in different social contexts (alone, single gender group, mixed gender group etc.). Furthermore, (iv) the ground truth quality of the dataset could be questionable.

3 Conclusion

We presented four scientific contributions for the application of affective recommender systems for multimedia items. We provided the underlying database acquisition and basic statistics, experimental results for the CBR and CF recommenders and a workplan for the emotion detection part.

Acknowledgement

We would like to thank the members of the LDOS group for their assistance in the dataset acquisition procedure as well as the students who participated in the experiment. A special thank goes to our colleagues Tomaž Požrl and Matevž Kunaver.

References

Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749.

Berger, H., Denk, M., Dittenbach, M., Pesenhofer, A., and Merkl, D. (2007). Photo-based user profiling for tourism recommender systems. In Psaila, G. and RolandWagner, editors, *E-Commerce and Web Technologies*, volume 4655/2007, pages 46–55. Springer Berlin / Heidelberg.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370.

Chibelushi, C. and Bourel, F. (2003). Facial expression recognition: A brief tutorial overview. *CVonline: On-Line Compendium of Computer Vision*, 9.

Cowie, R., Cowie, E. D., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1):32–80.

Donato, G., Bartlett, M. S., Hager, J. C., and Sejnowski, P. E. T. J. (1999). Classifying facial actions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(10):974–989.

Ekman, P. and Friesen, W. V. (2003). *Unmasking the face*. Malor Books, Cambridge MA 02238.

Fasel, B. and Luettin, J. (2003). Automatic facial expression analysis: A survey. *Pattern Recognition*, 35(2):259–275.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., and Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40:84–96.

González, G., López, B., and J. LL, d. l. R. (2004). Managing emotions in smart user models for recommender systems. In *Proceedings of 6th International Conference on Enterprise Information Systems ICEIS 2004*, volume 5, pages 187–194.

Hanjalic, A. (2006). Extracting moods from pictures and sounds. *IEEE Signal Processing Magazine*.

Herlocker, J. L., Konnstan, J. A., Terveen, L. G., and Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactionson Information Systems*, 22(1).

http://ipip.ori.org/ (last accessed: April 2009). International personality item pool: A scientific collaboratory for the development of advanced measures of personality and other individual differences. web site.

Kim, J. and André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12):2067–2083.

Kunaver, M., Požrl, T., Pogačnik, M., and Tasič, J. (2007). Optimisation of combined collaborative recommended systems. *International Journal of Electronic Communications*, 61:433–443.

Lang, P., Bradley., M., and Cuthbert, B. (2005). International affective picture system (iaps): Affective ratings of pictures and instruction manual. technical report a-6. Technical report, University of Florida, Gainesville, FL.

Lekakos, G. and Giaglis, G. M. (2006). Improving the prediction accuracy of recommendation algorithms: Approaches anchored on human factors. *Interacting with computers*, 18:410–431.

Lew, M. S., Sebe, N., Djeraba, C., and Jain, R. (2006). Content-based multimedia information rtrieval: State of the art and challenges. *ACM Transactions on multimedia computing*, 2(1):1–19.

Manjunath, B., Salembier, P., and Sikora, T., editors (2002). *Introduction to MPEG-7: Multimedia Content Description Interface*. Wiley.

Picard, R. and Daily, S. B. (2005). Evaluating affective interactions: Alternatives to asking what users feel. In *CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches*, Portland, OR.

Picard, R. W. (2000). *Affective Computing*. MIT Press.

Pogačnik, M., Tasič, J., Meža, M., and Košir, A. (2005). Personal content recommender based on a hierarchical user model for the selection of tv programmes. *User Modeling and User Adapted Interaction*, 15:425–457.

Posner, J., Russell, J. A., and Peterson, B. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17:715–734.

Schröder, M., Baggia, P., Burkhardt, F., Martin, J.-C., Pelachaud, C., Peter, C., Schuller, B., Wilson, I., and Zovato, E. (2008). Elements of an emotionml 1.0. W3C Incubator Group Report 20 November 2008.

Shan, M.-K., Kuo, F.-F., Chiang, M.-F., and Lee, S.-Y. (2009). Emotion-based music recommendation by affinity discovery from film music. *Expert Syst. Appl.*, 36(4):7666–7674.

Tkalčič, M., Tasič, J., and Košir, A. (2009). Usage of affective parameters in a content-based multimedia recommender system. *Submitted to the Elsevier International Journal of Human-Computer Studies*.

Villon, O. and Lisetti, C. (2006). A user-modeling approach to build user's psycho-physiological maps of emotions using bio-sensors. In *The 15th IEEE International Symposium on Robot and Human Interactive Communication, 2006. ROMAN 2006.*, pages 269–276.

Viola, P. and Jones, M. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154.

Wang, Y. and Guan, L. (2008). Recognizing human emotional state from audiovisual signals. *IEEE Transactionson multimedia*, 10(5):936–946.

Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(1):39–58.