ELEKTROTEHNIŠKI VESTNIK 78(1-2): 12–17, 2011
ENGLISH EDITION

Usage of affective computing in recommender systems

Marko Tkalčič, Andrej Košir, Jurij Tasič

University of Ljubljana, Faculty of Electrical Engineering, Trzaska 25, 1000 Ljubljana, Slovenia
E-mail: marko.tkalcic@fe.uni-lj.si

Abstract. In this paper we present the results of three investigations of our broad research on the usage of affect and personality in recommender systems. We improved the accuracy of a content-based recommender system with the inclusion of affective parameters in user and item modeling. We improved the accuracy of a content filtering recommender system under the cold start conditions with the introduction of a personality-based user similarity measure. Furthermore we developed a system for implicit tagging of images with affective metadata.

Key words: recommender system, affective computing, machine learning, user profile, emotions

1 INTRODUCTION

End users are having difficulties in finding relevant content items in today's large multimedia (MM) databases. Recommender systems (RS) help the end users by narrowing their choice of the MM content based on the users' preferences stored in user profiles. Today's RS do not meet the users' needs as the list of recommended items is not accurate [10]. The goal of this paper is to introduce affective computing [12] based methods for the improvement of recommender systems. We present a content-based recommender system based on affective metadata, a collaborative filtering recommender system based on personality metadata and a method for the affective labeling of images based on an emotion detection algorithm.

1.1 Problem statement

In order to increase the accuracy of RS there are two possible ways of doing it: (i) by optimizing the existing algorithms or (ii) by finding better features that describe more of the unexplained variance [8]. In this paper we present the improvement of the RS accuracy by introducing features based on the users' emotive responses and their personalities. These features describe a large part of the unexplained variance in the users' preferences. The preferences are expressed as ratings (e.g. Likert ratings, binary ratings etc.). The ratings can be acquired explicitly or implicitly (e.g. viewing time as an indicator of the rating [7]).

In our effort to improve the RS's accuracy we addressed the following three areas: (i) the usage of affective parameters in a content-based recommender (CBR) system, (i) an algorithm for the unobtrusive acquisition of affective labels and (iii) the usage of a personality based user similarity measure (USM) in a collaborative filtering (CF) RS. Fig. 1 shows the architecture of the affective recommender system and the places where we introduced the described advances.

The remainder of the paper is organised as follows. In section 2 we describe the dataset acquisition procedure. Sections 3, 4 and 5 present the three main contributions: a CBR recommender with affective metadata, a CF recommender with a personality-based USM and an emotion detection algorithm, respectively. Each of these sections contains the experiment description and results subsections. In section 6 we conclude the paper.

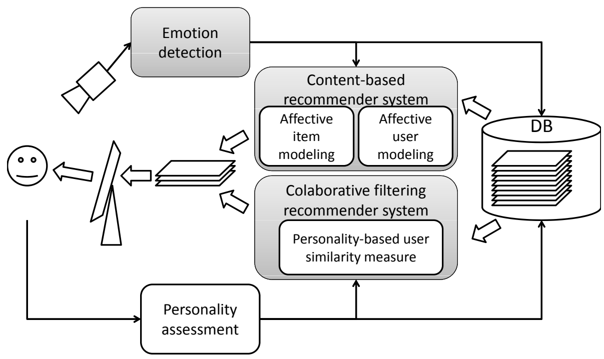

Figure 1. Basic architecture of the affective recommender system

1.2 Related work

The most common division of RS differs between (i) content based recommender systems, (ii) collaborative filtering recommender systems and (iii) hybrid recommender (HR) systems [1]. There are lots of recommender systems that use one of the generic user modeling approaches [1]. With the exception of the system developed by Arapakis [2] and Tkalčič [14] there have been no related work in recommender systems with

Received 13 October 2010
Accepted 1 February 2011

affective user modeling. Pantić and Vinciarelli suggest to use affective labels for tagging the content by using unobtrusive emotion detection techniques [11]. There are several recommender systems that use a generic approach in user modeling and are surveyed in [1]. Several investigations focus on the unobtrusive acquisition of users' emotions through various modalities [17]. The usage of affective metadata was already foreseen in the TVAnytime standard [13].

2 Dataset acquisition

In order to validate our hypotheses experimentally, we needed a suitable dataset. Despite the existence of several datasets with affective labels [6, 5] these do not include recommender systems related data, thus we had to build our own dataset. The dataset had to meet the following requirements: (i) a set of MM content with generic and affective labels, (ii) a set of users with related personality profiles in the form of the five-factor model (FFM [9]), (iii) video sequences of users' during the consumption of the MM content and (iv) explicit ratings.

We chose a subset of images from the IAPS dataset [3] as content items. These images, which were labeled with affective metadata in the valence-arousal-dominance (VAD) space, had two roles: content items and emotional stimuli. We chose 70 images from the IAPS dataset and we annotated them manually with the genre metadata. We built a computer application that showed the user an image that caused an emotive response that was recorded with a web camera placed above the monitor. The user also gave each image an explicit five-scale Likert rating. We had 52 subjects taking part in this experiment, called the emotion induction experiment [4], who also filled-in the IPIP personality questionnaire (http://ipip.ori.org) for the assessment of the FFM values. Thus we obtained, for each user, a five tuple describing the five personality factors (*Agreeableness, Conscientiousness, Openness, Neuroticism, Extraversion*).

Our dataset is named LDOS-PerAff-1 [15] and is accessible via web: http://slavnik.fe.uni-lj.si/PerAff.

3 A CBR system with affective metadata

The content items (e.g. films, music, images) used in content-based recommender systems are annotated with metadata (e.g. the genre, actors, theme etc.). We use h as the notation for the content item. The data structure, associated with the item h, also called the *item profile*, is denoted with $md(h)$. A CBR system filters out a limited set of content items based on the user's preferences towards specific metadata values. These preferences are stored in a data structure called the *user profile* $up(u)$ of user u. A good choice of metadata for the user profiles and the item profiles is of crucial importance in the design of CBR systems. These metadata must explain a large part of the variance for differentiating the relevant items from the non-relevant items for each individual user. We denote the explicit ratings given by the user u to the item h with $e(u,h)$ and they represent the ground truth for assessing the relevant and the non-relevant items.

We propose to use affective metadata to separate the relevant items the from non-relevant. We assume that the affective metadata explain more variance than generic metadata used in state of the art systems. This assumption is reflected by the fact that people have different tastes about the sought emotive state, as is the case of the paintings in Fig. 2. The paintings induce different emotive states. We verified this hypothesis by building a CBR system for images annotated with the affective metadata.

a. E. Munch b. C. Monet

Figure 2. Two paintings eliciting different emotions.

We compared the quality of the recommended items based on the generic metadata (GM) and the affective metadata (AM). GM consisted of the genre g and the average watching time $\bar{t_w}$. AM consisted of the mean and standard deviation of the emotive responses of users to the observed item.

There are several possibilities for describing an emotive response: as basic emotions, using the dimensional model or the circumplex model [14]. The most commonly used approach are the six basic emotions: joy, anger, fear, disgust, sadness and surprise. In the dimensional model, each emotion is described with the triple *valence*, *arousal* and *dominance* (VAD). The circumplex model maps the basic emotions into the dimensional model by assigning each basic emotion an area in the dimensional space, most commonly in the valence-arousal plane.

We used the dimensional model to describe the emotive response $er(u,h) = (v,a,d)$ of the user u while consuming the item h. The values (v,a,d) represent the dimensions valence, arousal and dominance. The set of users that has consumed the item h is denoted with U_h and their emotive responses form the set $ER_h = \{er(u,h) : u \in U_h\}$. We propose the item profile as the first two statistical moments of each dimension v, a and d which yields the six-tuple

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

We modeled the users using the machine learning (ML) algorithms described in the next subsection. Each user profile $up(u)$ was a data structure containing the parameters of the trained ML algorithm. Thus the data structure of the user models heavily depends on the ML algorithm used.

Based on the trained user model, we used the respective ML technique to estimate the ratings for the items that were not rated by the user. We denoted these ratings with $\hat{e}(u, h)$. Both the explicit ratings $e(u, h)$ and the estimated ratings $\hat{e}(u, h)$ can take only one of the two possible values, C_0 and C_1, which represent the classes of relevant and non-relevant items, respectively.

3.1 *Experiment*

The experiment was meant to prove the hypothesis that the quality of the recommended items can be improved by using AM. For this reason we ran the CBR simulation two times: once with AM and once with GM. In both cases we compared the estimated values with the ground truth explicit ratings using the ten-fold cross validation method. This procedure yielded the confusion matrix from which we computed the scalar measures precision P, recall R and F-measure. We also evaluated four different ML techniques: Naive Bayes, AdaBoost, C4.5 and Support Vector Machines (SVM). We used the Matlab and Weka (http://www.cs.waikato.ac.nz/ml/weka/) tools.

3.2 *Results*

The results are shown in Tab. 1. We performed the Pearson χ^2 statistical test for the equivalence of the confusion matrix distributions. The results showed that the confusion matrices of the AM and GM based CBR are significantly different. We further observed the scalar measures P, R and F and concluded that the CBR with AM performs significantly better than the same CBR with GM only. We also assessed the quality of the AM features and noted that the valence mean ($\bar{v}$) is the feature that carries the largest part of the output variance. The best ML technique turned out to be the SVM.

metadata	classifier γ	P	R	F
$\mathcal{A}$	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
$\mathcal{A} \times \mathcal{V}$	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
$\mathcal{V}$	AdaBoost	0.64	0.56	0.60
	C4.5	0.62	0.54	0.58
	NaiveBayes	0.57	0.60	0.58
	SVM	0.68	0.55	0.61

Table 1. P, R and F measures for different metadata sets and classifiers. The set $\mathcal{V}$ is composed of the mean and standard deviation of the VAD values.

4 A PERSONALITY-BASED USER SIMILARITY MEASURE

Collaborative filtering recommender systems suggest to the observed user those items that have been liked by the user's neighbours. The key element of such system is the user similarity measure (USM). Most systems calculate the similarity between users based on the past ratings of the same items. The users who have rated the same items in a similar way are thought as neighbours. The problem arises when a new user joins the system and there are not enough ratings to be used to assess the neighbours. This problem is called the *cold-start problem*. In order to alleviate it we propose to use a personality-based USM.

The users' personality is usually described with the Five Factor Model (FFM), also called the big5 model. In this model each user's personality is described with five scalar values reflecting the five basic personality factors: openness, conscientiousness, extraversion, agreeableness, and neuroticism. These factors account for the majority of variance in end users [9]. The underlying assumption for the choice of this approach is that users with similar personalities have similar preferences for multimedia content.

4.1 *Experiment*

We performed the simulation of the cold-start circumstances by taking into account only the first s rating of the observed users where we gradually increased s from 1 to the number of all ratings J. Then we switched to the personality based USM and ran again the CF recommender system.

Based on the USM employed, we determined the k nearest neighbours. We used $k = 7$. The equation was then applied to calculate the estimated ratings $\hat{e}(u, h)$.

$$\hat{e}(u, h) = \alpha \bar{e}^{NN}(u, h) + (1 - \alpha)\bar{e}^{P}(u, h) \quad (1)$$

As in the case of the CBR system, we calculated the confusion matrix for each user and then

The proposed USM was constructed using the personality values vectors for each user $\vec{b} = (b_1, \ldots, b_5)$, where the scalar values of the vector $b_l \in [0, 1]$ represent the five personality factors. We used the IPIP questionnaire (described in section 2) to assess the values of the vector $\vec{b}$. Then we used a weighted Euclidian distance to calculate the similarity between the users u_i and u_j represented by their respective personality vectors $\vec{b_i}$ and $\vec{b_j}$

$$d_W(\vec{b_i}, \vec{b_j}) = \sqrt{\sum_{l=1}^{5} w_l (b_{il} - b_{jl})^2} \tag{2}$$

4.2 Results

The comparison of F measures, which is shown in Fig. 3, shows that the quality of the recommended items in cold-start circumstances is significantly higher when using the proposed USM than the rating based USM when $s < 50$. This means that the personality based USM does alleviate the cold start problem as hypothesized. The main drawback of the proposed method is the annoying acquisition of the FFM values besides the ethical issues as personalities are very sensible information.

5 THE EMOTION DETECTION ALGORITHM FROM VIDEO SEQUENCES OF FACIAL EXPRESSIONS

The CBR system presented in this paper requires an implicit, unobtrusive method for tagging the content with affective metadata. Although emotion detection can use several modalities (e.g., voice, facial expressions, posture etc. [17]) we propose to use an emotion detection algorithm based on video sequences of facial expressions.

Emotion detection in related work is usually done on two types of datasets: posed and spontaneous. In the posed datasets, actors play facial expressions to their maximal extent. In the spontaneous datasets, an emotion induction technique is used to elicit emotive expressions in end users.

5.1 Experiment

We used an emotion induction technique for eliciting emotive responses in our users. We used 72 images from the IAPS dataset as emotion stimuli where each image was annotated with the induced emotions in the VAD space. These annotations were used as the ground truth in the training and test phases. The users' facial expressions were recorded with a webcam.

We compared the performance of the used emotion detection algorithm on our spontaneous LDOS-PerAff-1 dataset and on the posed Kanade-Cohn dataset.

We split the VAD space into eight subspaces by dividing each axis (V, A and D) in two equal parts. Among these eight new classes in the VAD space, two did not have any samples in, so we performed a six-class emotion detection.

We first segmented the video sequences into shorter clips of emotion responses to single stimuli. Using the Viola-Jones algorithm [16], we extracted the user's face, which we registered and normalized (see Fig. 4).

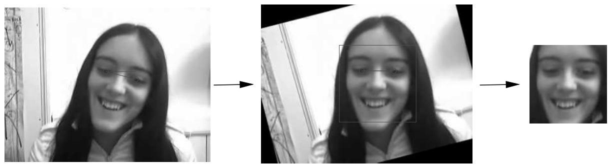

Figure 4. Preprocessing of face frames.

We compared the neutral face frame with the maximum expression face frame. In the Kanade-Cohn dataset, we used the first and the last frame. In the LDOS-PerAff-1 dataset, we calculated the neutral frame by averaging all the frames of a single user.

Then we extracted the low-level features from both frames using the Gabor filtering. We aggregated the features from both frames using absolute values, differences and quotients of the both frames' features. We used the kNN classifier for training and detection. We split the dataset into the training and test set using two thirds as the training set and the remaining third as the test set. We assessed the quality of the emotion detection algorithm using the scalar measures derived from the confusion matrix.

5.2 Results

The accuracy achieved on the posed dataset was on the level of 92% and on the spontaneous dataset 62%. Tab. 2 shows the confusion matrix. The obtained results are in line with similar investigations.

classified as	1	2	3	4	5	6
1	15	3	0	4	5	6
2	4	456	18	19	107	48
3	0	57	98	10	51	29
4	2	42	3	156	41	13
5	2	101	19	28	477	64
6	1	82	11	21	104	240

Table 2. Confusion matrix for the six emotive classes of the dataset of spontaneous facial expressions LDOS-PerAff-1.

We can conclude that the presented method for the detection of the users' emotive states from video sequences of the users' facial expressions is not suitable for unobtrusive affective tagging of images and for being used in a content-based recommender system

6 CONCLUSION

The presented new approaches toward the improvement of the RS accuracy show positive results. The usage of the affective metadata in the CBR system yield significantly better accuracy than the usage of the generic metadata. The proposed personality based user similarity measure provides significantly better accuracy than the rating based USM under the cold start conditions in the CF recommender system. Unfortunately, the algorithm for the detection of emotions in spontaneous videoclips

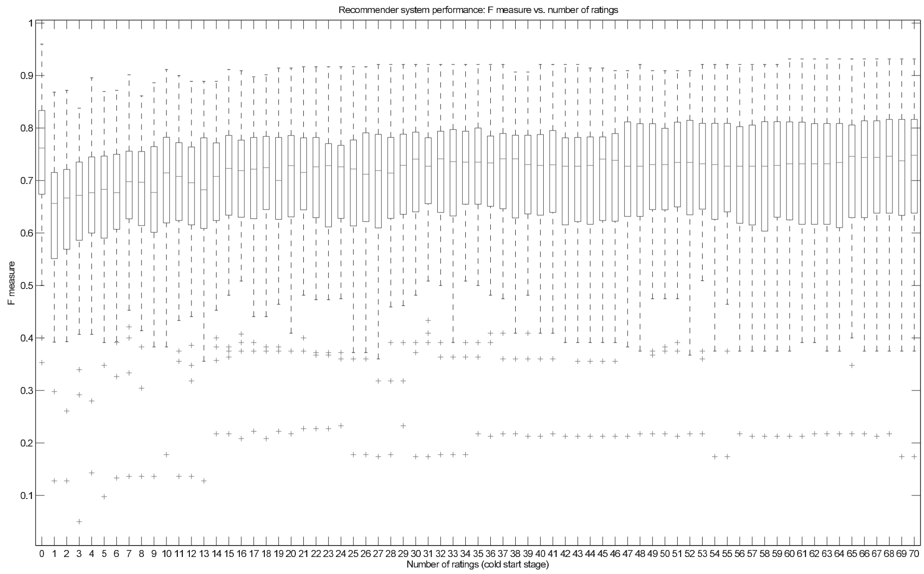

Figure 3. Boxplots of F values. The leftmost boxplot (at $s = 0$) shows the results of the proposed user similarity measure. Each boxplot shows the distribution of the F values for different users at a fixed stage s. The box contains the values between the 25th and 75th percentile of the F measures.

is not ready yet for the labeling of content items with affective tags.

In order to use affective recommender systems in real applications, the algorithms for implicit affective tagging of MM content should be improved.

ACKNOWLEDGEMENTS

The work has been supported by the Slovenian research agency http://arrs.gov.si under the contract P2-0246. The authors would like to thank the students and the staff at the Gimnazija Poljane school for their involvement in the dataset acquisition. Furthermore, the authors are thankful to all the colleagues from the LDOS group that helped in the experimental part.

REFERENCES

[1] G. Adomavicius and A. Tuzhilin. 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, 2005.

[2] I. Arapakis, Y. Moshfeghi, H. Joho, R. Ren, D. Hannah, J.M. Jose, and L. Gardens. Integrating facial expressions into user profiling for the improvement of a multimodal recommender system. In *Proc. IEEE Int'l Conf. Multimedia & Expo*, pages 1440–1443, 2009.

[3] Margaret M. Bradley and Lang. *The International Affective Picture System (IAPS) in the Study of Emotion and Attention*, chapter 2. Series in Affective Science. Oxford University Press, 198 Madison Avenue, 2007.

[4] J.A. Coan and J.J.B. Allen. *Handbook of emotion elicitation and assessment*. Oxford university press, 2007.

[5] R. Gajšek, V. Štruc, F. Mihelič, A. Podlesek, L. Komidar, G. Socan, B. Bajec, and P. Methodology. Multi-Modal Emotional Database: AvID. *Informatica*, 33:101–106, 2009.

[6] T. Kanade, J.F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the fourth IEEE International conference on Automatic Face and Gesture Recognition*, page 46, 2000.

[7] Y.S. Kim, B.J. Yum, J. Song, and S.M. Kim. Development of a recommender system based on navigational and behavioral patterns of customers in e-commerce sites. *Expert Systems with Applications*, 28(2):381–393, 2005.

[8] Michael S. Lew, Nicu Sebe, Chabane Djeraba, and Ramesh Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on multimedia computing*, 2(1):1–19, February 2006.

[9] Robert R. McCrae and Oliver P. John. An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2):p175 – 215, 06 1992.

[10] Leavitt Neal. Recommendation technology: Will it boost e-commerce. *IEEE Computer*, 39(5):13–16, May 2006.

[11] M. Pantic and A. Vinciarelli. Implicit Human-Centered Tagging. *IEEE Signal Processing Magazine*, 26(6):173–180, 2009.

[12] Rosalind W. Picard. *Affective Computing*. MIT Press, 2000.

[13] Skip Pizzi, Ronald Tol, Henry Chadwick, and Masahito Kawamori. Tvanytime system description (informative with normative appendix b). Technical report, TV Anytime, 2003.

[14] Marko Tkalčič, Urban Burnik, and Andrej Košir. Using affective parameters in a content-based recommender system. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, 20(4):?–?5, 2010. ?

[15] Marko Tkalčič, Jurij Tasič, and Andrej Košir. The LDOS-

PerAff-1 Corpus of Face Video Clips with Affective and Personality Metadata. In Michael Kipp, editor, *Proceedings of the LREC 2010 Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, 2010.

[16] Paul Viola and Michael Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[17] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(1):39–58, Jan. 2009.

Marko Tkalčič is a research scientist at the University of Ljubljana Faculty of electrical engineering. His research areas include affective computing, user modeling, recommender systems and human computer interaction.

Andrej Košir is an associate professor at the University of Ljubljana Faculty of electrical engineering. His research areas include engaged in the research of digital image and image understanding, user modeling and personalization, statistical data modeling and operational research in telecommunications.

Jurij Tasič received his B.Sc,, M.Sc and PhD in electrical engineering from University of Ljubljana in 1971, 1973 and 1977 respectively. He conducts teaching and researches in the fields of signal and image processing, telecommunication systems, adaptive algorithms and parallel algorithms for many years. His current interests are in the Advanced Algorithms in Signal and Image processing, Communication systems, Parallel Algorithms and new wideband services at high speed gigabit systems.