Impact of implicit and explicit affective labeling on a recommender system's performance

Marko Tkalčič, Ante Odić, Andrej Košir, and Jurij Tasič

University of Ljubljana Faculty of electrical engineering,
Tržaška 25, 1000 Ljubljana, Sovenia
{marko.tkalcic,ante.odic,andrej.kosir,jurij.tasic}@fe.uni-lj.si
http://ldos.fe.uni-lj.si

Abstract. Affective labeling of multimedia content can be useful in recommender systems. In this paper we compare the effect of implicit and explicit affective labeling in an image recommender system. The implicit affective labeling method is based on an emotion detection technique that takes as input the video sequences of the users' facial expressions. It extracts Gabor low level features from the video frames and employs a kNN machine learning technique to generate affective labels in the valence-arousal-dominance space. We performed a comparative study of the performance of a content-based recommender (CBR) system for images that uses three types of metadata to model the users and the items: (i) generic metadata, (ii) explicitly acquired affective labels and (iii) implicitly acquired affective labels with the proposed methodology. The results showed that the CBR performs best when explicit labels are used. However, implicitly acquired labels yield a significantly better performance of the CBR than generic metadata while being an unobtrusive feedback tool.

Keywords: content-based recommender system, affective labeling, emotion detection, facial expressions, affective user modeling

1 Introduction

Recently, investigations, that evaluate the use of *affective metadata* (AM) in content-based recommender (CBR) systems, were carried out [Arapakis et al., 2009, Tkalčič et al., 2010a] and showed an increase of the accuracy of recommended items. This improvement of CBR systems that use affective metadata over systems that use *generic metadata* (GM), like the genre, represents the motivation for the work presented in this paper. Such systems require that the content items are labeled with affective metadata which can be done in two ways: (i) explicitly (i.e. asking the user to give an explicit affective label for the observed item) or (ii) implicitly (i.e. automatically detecting the user's emotive response).

1.1   Problem statement and proposed solution

Each of the two approaches for affective labeling, explicit and implicit, has its pros and cons. The explicit approach provides unambiguous labels but Pantic and Vinciarelli [2009] argue that the truthfulness of such labels is questionable as users can be driven by different motives (egoistic labeling, reputation-driven labeling and asocial labeling). Another drawback of the explicit labeling approach is the intrusiveness of the process. On the other hand implicit affective labeling is completely unobtrusive and harder to be cheated by the user. Unfortunately the accuracy of the algorithms that detect affective responses might be too low and thus yield ambiguous/inaccurate labels.

Given the advantages of implicit labeling over explicit there is a need to assess the impact of the low emotion detection accuracy on the performance of recommender systems.

In this paper we compare the performance of a CBR system using explicit affective labeling vs. the proposed implicit affective labeling. The baseline results of the CBR with explicit affective labeling are those published in Tkalčič et al. [2010a]. The comparative results of the implicit affective labeling are obtained using the same CBR procedure as in Tkalčič et al. [2010a], the same user interaction dataset [Tkalčič et al., 2010c] but with affective labels acquired implicitly.

1.2   Related work

As anticipated by Pantic and Vinciarelli [2009], affective labels are supposed to be useful in content retrieval applications. Work related to this paper is divided in (i) the acquisition of affective labels and (ii) the usage of affective labels.

The acquisition of explicit affective labels is usually performed through an application with a graphical user interface (GUI) where users consume the multimedia content and provide appropriate labels. An example of such an application is the one developed by Eckhardt and Picard [2009].

On the other hand, the acquisition of implicit affective labels is usually reduced to the problem of non-intrusive emotion detection. Various modalities are used, such as video of users' faces, voice or physiological sensors (heartbeat, galvanic skin response etc.) [Picard and Daily, 2005]. A good overview of such methods is given in Zeng et al. [2009]. In our work we use implicit affective labeling from videos of users' faces. Generally, the approach taken in related work in automatic detection of emotions from video clips of users' faces is composed of three stages: (i) pre-processing, (ii) low level features extraction and (iii) classification. Related work differ mostly in the last two stages. Bartlett et al. [2006], Wang and Guan [2008], Zhi and Ruan [2008] used Gabor wavelets based features for emotion detection. Beside these, which are mostly used, Zhi and Ruan [2008] report the usage of other facial features in related work: active appearance models (AAM), action units, various facial points and motion units, Haar based features and textures. Various classification schemes were used successfully in

video emotion detection. Bartlett et al. [2006] employed both the Support Vector Machine (SVM) and AdaBoost classifiers. Zhi and Ruan [2008] used the k-nearest neighbours (k-NN) algorithm. Before using the classifier they performed a dimensionality reduction step using the locality preserving projection (LPP) technique. In their work, Wang and Guan [2008] compared four classifiers: the Gaussian Mixture Model (GMM), the k-NN, neural networks (NN) and Fisher's Linear Discriminant Analysis (FLDA). The latter turned out to yield the best performance. The survey Zeng et al. [2009] reports the use of other classifiers like the C4.5, Bayes Net and rule based classifiers. Joho et al. [2009] used an emotion detection techique that uses video sequences of users' face expressions to provide affective labels for video content.

Another approach is to extract affective labels directly from the content itself, without observing the users. Hanjalic and Xu [2005] used low level features extracted from the audio track of video clips to identify moments in video sequences that induce high arousal in viewers.

In contrast to emotion detection techniques the usage of affective labels for information retrieval has only recently started to gain attention. Chen et al. [2008] developed the EmoPlayer which has a similar user interface to the tool developed by Eckhardt and Picard [2009] but with a reversed functionality: it assists users to find specific scenes in a video sequence. Soleymani et al. [2009] built a collaborative filtering system that retrieves video clips based on affective queries. Similarly, but for music content, Shan et al. [2009] have developed a system that performs emotion based queries. Arapakis et al. [2009] built a complete video recommender system that detects the users' affective state and provides recommended content. Kierkels and Pun [2009] used physiological sensors (ECG and EEG) to implicitly detect the emotive responses of users. Based on implicit affective labels they observed an increase of content retrieval accuracy compared to explicit affective labels. Tkalčič et al. [2010a] have shown that the usage of affective labels significantly improves the performance of a recommender system over generic labels.

2 Affective modeling in CBR systems

2.1 Emotions during multimedia items consumption

In a multimedia consumption scenario a user is watching multimedia content. During the consumption of multimedia content (images in our case), the emotive state of a user is continuously changing between different emotive states $\epsilon_j \in E$, as different visual stimuli $h_i \in H$ induce these emotions (see Fig. 1). The facial expressions of the user are being continuously monitored by a video camera for the purpose of the automatic detection of the emotion expressions.

The detected emotion expressions of the users, along with the ratings given to the content items, can be used in two ways: (i) to model the multimedia content item (e.g. *the multimedia item h_i is funny - it induces laughter in most of the viewers*) and (ii) to model individual users (e.g. *the user u likes images that induce fear*).
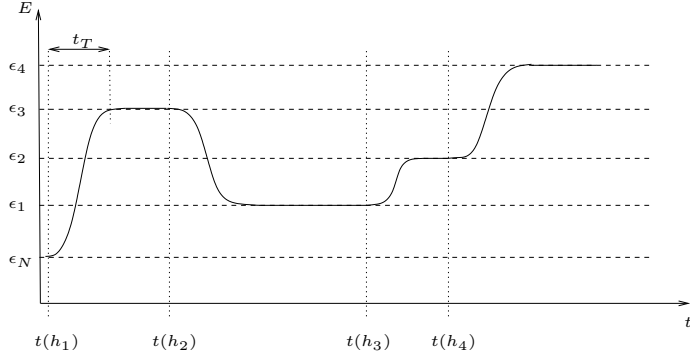
Fig. 1: The user's emotional state ϵ is continuously changing as the the time sequence of the visual stimuli $h_i \in H$ induce different emotions.

2.2 Affective modeling in a CBR system

Item modeling with affective metadata We use the valence-arousal-dominance (VAD) emotive space for describing the users' emotive reactions to images. In the VAD space each emotive state is described by three parameters, namely valence, arousal and dominance. A single user $u \in U$ consumes one or more content items (images) $h \in H$. As a consequence of the image h being a visual stimulus, the user u experiences an emotive response which we denote as $er(u, h) = (v, a, d)$ where v, a and d are scalar values that represent the valence, arousal and dominance dimensions of the emotive response er. The set of users that have watched a single item h are denoted with U_h. The emotive responses of all users U_h, that have watched the item h form the set $ER_h = \{er(u, h) : u \in U_h\}$. We model the image h with the item profile that is composed of the first two statistical moments of the VAD values from the emotive responses ER_h which yields the six tuple

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

where $\bar{v}$, $\bar{a}$ and $\bar{d}$ represent the average VAD values and σ_v, σ_a and σ_d represent the standard deviations of the VAD values for the observed content item h. An example of the affective item profile is shown in Tab. 1.

User modeling with affective metadata The preferences of the user are modeled based on the explicit ratings that she/he has given to the consumed items. The observed user u rates each viewed item either as relevant or non-relevant. A machine learning (ML) algorithm is trained to separate relevant from non-relevant items using the affective metadata in the item profiles as features and the binary ratings (relevant/non-relevant) as classes. The user profile $up(u)$ of the observed user u is thus an ML algorithm dependent data structure. Fig. 2 shows an example of a user profile when the tree classifier C4.5 is being used.

Metadata field	Value
$\bar{v}$	3.12
σ_v	1.13
$\bar{a}$	4.76
σ_a	0.34
$\bar{d}$	6.28
σ_d	1.31

Table 1: Example of an affective item profile $\mathcal{V}$ (first two statistical moments of the induced emotion values v, a and d) .
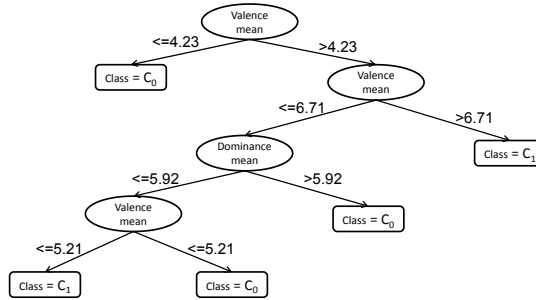

Fig. 2: Example of a user profile when the C4.5 tree classifier is used for inferring the user's preferences. The labels C_0 and C_1 represent the relevant and non-relevant classes, respectively.

3 Experiment

We used our implementation of an emotion detection algorithm (see Tkalčič et al. [2010b]) for implicit affective labeling and we compared the performance of the CBR system that uses explicit vs. implicit affective labels.

3.1 Overview of the emotion detection algorithm for implicit affective labeling

The emotion detection procedure used to give affective labels to the content images involved three stages: (i) pre-processing, (ii) low level feature extraction and (iii) emotion detection. We formalized the procedure with the mappings

$$I \rightarrow \Psi \rightarrow E \tag{2}$$

where I represents the frame from the video stream, Ψ represents the low level features corresponding to the frame I and E represents the emotion corresponding to the frame I.

In the pre-processing stage we extracted and registered the faces from the video frames to allow precise low level feature extraction. We used the eye tracker developed by Valenti et al. [2009] to extract the locations of the eyes. The detection of emotions from frames in a video stream was performed by comparing the current video frame I_t of the user's face to a neutral face expression. As the LDOS-PerAff-1 database is an ongoing video stream of users consuming different images we averaged all the frames to get the neutral frame. This method is applicable when we have a non supervised video stream of a user with different face expressions.

The low level features used in the proposed method were drawn from the images filtered by a Gabor filter bank. We used a bank of Gabor filters of 6 different orientation and 4 different spatial sub-bands which yielded a total of 24 Gabor filtered images per frame. The final feature vector had the total length of 240 elements.

The emotion detection was done by a k-NN algorithm after performing dimensionality reduction using the principal component analysis (PCA).

Each frame from the LDOS-PerAff-1 dataset was labeled with a six tuple of the induced emotion $\mathcal{V}$. The six tuple was composed of scalar values representing the first two statistical moments in the VAD space. However, for our purposes we opted for a coarser set of emotional classes $\epsilon \in E$. We divided the whole VAD space into 8 subspaces by thresholding each of the three first statistical moments $\bar{v}$, $\bar{a}$ and $\bar{d}$. We thus gained 8 rough classes. Among these, only 6 classes actually contained at least one item so we reduced the emotion detection problem to a classification into 6 distinct classes problem as shown in Tab. 2.

class E	$\bar{v}$	$\bar{a}$	$\bar{d}$	centroid values		
				v	a	d
ϵ_1	$\bar{v} > 0$	$\bar{a} < 0$	$\bar{d} < 0$	0.5	−0.5	−0.5
ϵ_2	$\bar{v} < 0$	$\bar{a} > 0$	$\bar{d} < 0$	−0.5	0.5	−0.5
ϵ_3	$\bar{v} > 0$	$\bar{a} > 0$	$\bar{d} < 0$	0.5	0.5	−0.5
ϵ_4	$\bar{v} < 0$	$\bar{a} < 0$	$\bar{d} > 0$	−0.5	−0.5	0.5
ϵ_5	$\bar{v} > 0$	$\bar{a} < 0$	$\bar{d} > 0$	0.5	−0.5	0.5
ϵ_6	$\bar{v} > 0$	$\bar{a} > 0$	$\bar{d} > 0$	0.5	0.5	0.5

Table 2: Division of the continuous VAD space into six distinct classes $E = \{\epsilon_1 \dots \epsilon_6\}$ with the respective centroid values.

3.2 Overview of the CBR procedure

Our scenario consisted in showing end users a set of still color images while observing their facial expressions with a camera. These videos were used for implicit affective labeling. The users were also asked to give explicit binary ratings to the images. They were instructed to select images for their computer wallpapers. The task of the recommender system was to select the relevant items for each user as accurate as possible. This task falls in the category *find all good items* for the recommender systems' tasks taxonomy proposed by Herlocker et al. [2004].

Figure 3 shows the overview of the CBR experimental setup. After we collected the ratings and calculated the affective labels for the item profiles, we trained the user profiles with four different machine learning algorithms: the SVM, NaiveBayes, AdaBoost and C4.5. We split the dataset in the train and test sets using the ten-fold cross validation technique. We then performed ten training/classifying iterations which yielded the confusion matrices that we used to assess the performance of the CBR system.
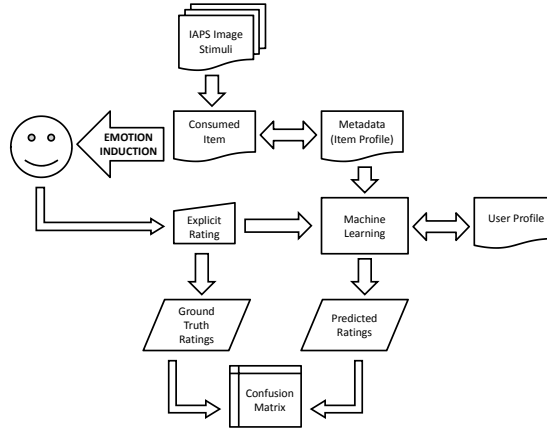

Fig. 3: Overview of the CBR experiment.

The set of images $h \in H$ that the users were consuming, had a twofold meaning: (i) they were used as content items and (ii) they were used as emotion induction stimuli for the affective labeling algorithm. We used a subset of 70 images from the IAPS dataset Lang et al. [2005]. The IAPS dataset of images is annotated with the mean and standard deviations of the emotion responses in the VAD space which was useful as the ground truth in the affective labeling part of the experiment.

The affective labeling algorithm described in Sec. 3.1 yielded rough classes in the VAD space. In order to build the affective item profiles we used the classes' centroid values (see Tab. 2) in the calculation of the first two statistical moments. We applied the procedure from Sec. 2.2.

We had 52 users taking part in our experiment (mean = 18.3 years, 15 males).

3.3 Affective CBR system evaluation methodology

The results of the CBR system were the confusion matrices of the classification procedure that mapped the images H into one of the two possible classes: relevant or non-relevant class. From the confusion matrices we calculated the recall, precision and F measure as defined in Herlocker et al. [2004].

We also compared the performances of the CBR system with three types of metadata: (i) generic metadata (genre and watching time as done by Tkalčič

et al. [2010a]), (ii) affective metadata given explicitly and (iii) affective meta-data acquired implicitly with the proposed emotion detection algorithm. For that purpose we transferred the statistical testing of the confusion matrices into the testing for the equivalence of two estimated discrete probability distributions [Lehman and Romano, 2005]. To test the equivalence of the underlying distributions we used the Pearson χ^2 test. In case of significant differences we used the scalar measures precision, recall and F measure to see which approach was significantly better.

4   Results

We compared the performance of the classification of items into relevant or non relevant through the confusion matrices in the following way: (i) Explicitly acquired affective metadata vs Implicitly acquired metadata, (ii) explicitly acquired metadata vs. generic metadata and (iii) implicitly acquired metadata vs. generic metadata. In all three cases the p value was $p < 0.01$. Table 3 shows the scalar measures precision, recall and F measures for all three approaches.

metadata/labeling method	classifier	P	R	F
implicit affective labeling	AdaBoost	0.61	0.57	0.59
	C4.5	0.58	0.50	0.53
	NaiveBayes	0.56	0.62	0.59
	SVM	0.64	0.47	0.54
explicit affective labeling	AdaBoost	0.64	0.56	0.60
	C4.5	0.62	0.54	0.58
	NaiveBayes	0.56	0.59	0.58
	SVM	0.68	0.54	0.60
generic metadata	AdaBoost	0.57	0.41	0.48
	C4.5	0.60	0.45	0.51
	NaiveBayes	0.58	0.57	0.58
	SVM	0.61	0.55	0.58

Table 3: The scalar measures P, R, F for the CBR system

5   Discussion

As we already reported in Tkalčič et al. [2010b], the application of the emotion detection algorithm on spontaneous face expression videos has a low performance. We identified three main reasons for that: (i) weak supervision in learning, (ii) non-optimal video acquisition and (iii) non-extreme facial expressions.

In supervised learning techniques there is ground truth reference data to which we compare our model. In the induced emotion experiment the ground

truth data is weak because we did not verify whether the emotive response of the user equals to the predicted induced emotive response.

Second, the acquisition of video of users' expressions in real applications takes place in less controlled environments. The users change their position during the session. This results in head orientation changes, size of the face changes and changes of camera focus. All these changes require a precise face tracker that allows for fine face registration. Further difficulties are brought by various face occlusions and changing lighting conditions (e.g. a light can be turned on or off, the position of the curtains can be changed etc.) which confuse the face tracker. It is important that the face registration is done in a precisely manner to allow the detection of changes in the same areas of the face.

The third reason why the accuracy drops is the fact that face expressions in spontaneous videos are less extreme than in posed videos. As a consequence the changes on the faces are less visible and are hidden in the overall noise of the face changes. The dynamics of face expressions depend on the emotion amplitude as well as on the subjects' individual differences.

The comparison of the performance of the CBR with explicit vs. implicit affective labeling shows significant differences regardless of the ML technique employed to predict the ratings. The explicit labeling yields superior CBR performance than the implicit labeling. However, another comparison, that between the implicitly acquired affective labels and generic metadata (genre and watching time) shows that the CBR with implicit affective labels is significantly better than the CBR with generic metadata only. Although not as good as explicit labeling, the presented implicit labeling technique brings additional value to the CBR system used.

The usage of affective labels is not present in state-of-the-art commercial recommender systems, to the best of the authors' knowledge. The presented approach allows to upgrade an existing CBR system by adding the unobtrusive video acquisition of users' emotive responses. The results showed that the inclusion of affective metadata, although acquired with a not-so-perfect emotion detection algorithm, significantly improves the quality of the selection of recommended items. In other words, although there is a lot of noise in the affective labels acquired with the proposed method, these labels still describe more variance in users' preferences than the generic metadata used in state-of-the-art recommender systems.

5.1 Pending issues and future work

The usage of affective labels in recommender systems has not reached a production level yet. There are several open issues that need to be addressed in the future.

The presented work was verified on a sample of 52 users of a narrow age and social segment and on 70 images as content items. The sample size is not big but it is in line with sample sizes used in related work [Arapakis et al., 2009, Joho et al., 2009, Kierkels and Pun, 2009]. Although we correctly used the statistical tests and verified the conditions before applying the tests a repetition of the

experiment on a larger sample of users and content items would increase the strength of the results reported.

Another aspect of the sample size issue is the impact of the size on the ML techniques used. The sample size in the emotion detection algorithm (the kNN classifier) is not problematic. It is, however, questionable the sample size used in the CBR. In the ten fold cross validation scheme we used 63 items for training the model and seven for testing. Although it appears that this is small, a comparison with other recommender system reveals that this is a common issue, and is usually referred as the sparsity problem. It occurs when, even if there are lots of users and lots of items, each user usually rated only few items and there are few data to build the models upon [Adomavicius and Tuzhilin, 2005].

The presented work also lacks a further user satisfaction study. Besides just aiming at the prediction of user ratings for unseen items research should also focus on the users' satisfaction with the list of recommended items.

But the most important thing to do in the future is to improve the emotion detection algorithms used for implicit affective labeling. In the ideal case, the perfect emotion detection algorithm would yield CBR performance that is identical to the CBR performance with explicit labeling.

The acquisition of video of users raises also privacy issues that need to be addressed before such a system can go in production.

Last, but not least, we believe that implicit affective labeling should be complemented with context modeling to provide better predictions of users' preferences. In fact, emotional responses of users and their tendencies to seek one kind of emotion over another, is tightly connected with the context where the items are consumed. Several investigations started to explore the influence of various contextual parameters, like being alone or being in company, on the users' preferences [Adomavicius et al., 2005, Odić et al., 2010]. We will include this information in our future affective user models.

6   Conclusion

We performed a comparative study of a CBR system for images that uses three types of metadata: (i) explicit affective labels, (ii) implicit affective labels and (iii) generic metadata. Although the results showed that the explicit labels yielded better recommendations than implicit labels, the proposed approach significantly improves the CBR performance over generic metadata. Because the approach is unobtrusive it is feasible to upgrade existing CBR systems with the proposed solution. The presented implicit labeling technique takes as input video sequences of users' facial expressions and yields affective labels in the VAD emotive space. We used Gabor filtering based low level features, PCA for dimensionality reduction and the kNN classifier for affective labeling.

Acknowledgement

References

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.

G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin. Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems (TOIS)*, 23(1):103–145, 2005.

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.

M.S. Bartlett, G.C. Littlewort, M.G. Frank, C. Lainscsek, I. Fasel, and J.R. Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, 1(6):22–35, 2006.

Ling Chen, Gen-Cai Chen, Cheng-Zhe Xu, Jack March, and Steve Benford. Emoplayer: A media player for video clips with affective annotations. *Interacting with Computers*, 20(1):17–28, January 2008. doi: http://dx.doi.org/10.1016/j.intcom.2007.06.003. URL `http://dx.doi.org/10.1016/j.intcom.2007.06.003`.

Micah Eckhardt and Rosalind Picard. A more effective way to label affective expressions. *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–2, September 2009. doi: 10.1109/ACII.2009.5349528. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5349528`.

Alan Hanjalic and Li-Qun Xu. Affective video content representation and modeling. *IEEE Transactions on Multimedia*, 7(1):143–154, February 2005.

J.L. Herlocker, J.A. Konstan, L.G. Terveen, and J.T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1):53, January 2004.

H. Joho, J.M. Jose, R. Valenti, and N. Sebe. Exploiting facial expressions for affective video summarisation. In *Proceeding of the ACM International Conference on Image and Video Retrieval*, pages 1–8. ACM, 2009.

J.J.M. Kierkels and T. Pun. Simultaneous exploitation of explicit and implicit tags in affect-based multimedia retrieval. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–6. IEEE, 2009.

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

E. L. Lehman and J.P. Romano. *Testing Statistical Hypotheses*. Springer Science + Business Inc., 2005.

Ante Odić, Matevž Kunaver, Jurij Tasič, and Andrej Košir. Open issues with contextual information in existing recommender system databases. *Proceedings of the IEEE ERK 2010*, A:217–220, September 2010.

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

Rosalind Picard and Shaundra Briant Daily. Evaluating affective interactions: Alternatives to asking what users feel. In *CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches*, Portland, OR, April 2005.

Man-Kwan Shan, Fang-Fei Kuo, Meng-Fen Chiang, and Suh-Yin Lee. Emotion-based music recommendation by affinity discovery from film music. *Expert Syst. Appl.*, 36(4):7666–7674, 2009. ISSN 0957-4174. doi: http://dx.doi.org/10.1016/j.eswa.2008.09.042.

Mohammad Soleymani, Jeremy Davis, and Thierry Pun. A collaborative personalized affective video retrieval system. *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–2, September 2009. doi: 10.1109/ACII.2009.5349526. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5349526`.

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), 2010a.

Marko Tkalčič, Ante Odić, Andrej Košir, and Jurij Tasič. Comparison of an emotion detection technique on posed and spontaneous datasets. *Proceedings of the IEEE ERK 2010*, 2010b.

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*, 2010c.

R. Valenti, Z. Yucel, and T. Gevers. Robustifying eye center localization by head pose cues. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009. URL `http://www.science.uva.nl/research/publications/2009/ValentiCVPR2009`.

Yongjin Wang and Ling Guan. Recognizing human emotional state from audio-visual signals. *IEEE Transactionson multimedia*, 10(5):936–946, 2008.

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. ISSN 0162-8828. doi: 10.1109/TPAMI.2008.52.

R. Zhi and Q. Ruan. Facial expression recognition based on two-dimensional discriminant locality preserving projections. *Neurocomputing*, 71(7-9):1730–1734, 2008.