

# Modeling Facial Expressions and Peripheral Physiological Signals to Predict Topical Relevance

Ioannis Arapakis, Ioannis Konstas,  
Joemon M. Jose  
Department of Computing Science  
University of Glasgow  
Glasgow, G12 8QQ  
{arapakis,konstas,jj}@dcs.gla.ac.uk

Ioannis Kompatsiaris  
Informatics and Telematics Institute  
Centre for Research and Technology Hellas  
Thermi-Thessaloniki, 57001  
ikom@iti.gr

## ABSTRACT

By analyzing explicit & implicit feedback information retrieval systems can determine topical relevance and tailor search criteria to the user's needs. In this paper we investigate whether it is possible to infer what is relevant by observing user affective behaviour. The sensory data employed range between facial expressions and peripheral physiological signals. We extract a set of features from the signals and analyze the data using classification methods, such as SVM and KNN. The results of our initial evaluation indicate that prediction of relevance is possible, to a certain extent, and implicit feedback models can benefit from taking into account user affective behavior.

**Categories and Subject Descriptors:** H.3.3 [Information Search and Retrieval]: Relevance Feedback, Search Process; I.5.1 [Computing Methodologies]: Pattern Recognition[Models]

**General Terms:** Experimentation, Human Factors, Performance

## 1. INTRODUCTION

To deal with information uncertainty, retrieval systems have employed in the past a range of relevance feedback techniques, which vary from explicit to implicit. Relevance is a key notion of information science, which is intertwined with many interactive processes, such as the act of communication, information seeking, assessment, reflection, etc. People very often apply relevance when performing some kind of information processing task, in order to determine the degree of appropriateness or relation of the perused information item [4].

By analyzing explicit and implicit feedback information the retrieval systems can determine topical relevance with greater accuracy, offer improved query reformulations and tailor the search criteria to the user needs. However, as discussed in [1], existing feedback techniques determine content relevance only with respect to the cognitive and situational levels of interaction, failing to acknowledge the importance of intentions, motivations and feelings in cognition and decision-making [3]. According to McKechnie and Ross [2], affective variables can play an important role in reading-related information behavior, especially in the domain of ev-

eryday life. Information processing, which can occur during the appraisal process of a goal, an event, or an item, often results in a series of changes in the user's cognitive and affective states.

We argue that such changes are often expressed through a psycho-physiological mobilization, which is reflected by a series of more or less observable cues (facial expressions, localized changes in the electrodermal activity, variations in the skin temperature, etc.). Since the significance of an event or information item can vary from low to high, depending on the number of goals or needs that are affected by it, so do these peripheral physiological symptoms vary in intensity and duration. We do not assume anything about the details of the relationship between users' affective responses and topical relevance; we systematically build our ground truth from the data, using classification and pattern recognition methods. This work is a study of how effectively we can determine topical relevance by measuring key physiological signals taken from the user. To justify our key assumption we make the following research hypothesis:

**H<sub>1</sub>:** Users' affective responses, as determined from automatic facial expression recognition, will vary across the relevancy of perused information items.

**H<sub>2</sub>:** Users' affective responses, as determined from peripheral physiological signal processing, will vary across the relevancy of perused information items.

## 2. EXPERIMENT

In this study we employed a facial expression recognition system (eMotion) [5], of reasonably robust performance and accuracy across all individuals. In addition, we recorded a range of peripheral physiological signals using a set of wearable devices (Polar RS800 and BodyMedia SenseWear<sup>®</sup> Pro<sub>3</sub> Armband), which gave us a more fine-grained imprint of participants' affective states. For the completion of the search tasks we used a customized version of Verge Engine [6], which worked on top of TREC 9 (2000) Web Track and TRECVID 2007 test collections. Twenty-four participants (14 male and 10 female) applied for the study through an organisational wide ad. Each participant completed four search tasks, two for each media type. In each task they were handed four topics and were asked to proceed with the one they considered more interesting. For every topic the system returned ten pre-selected results (either exclusively relevant or irrelevant), which the participants had to evaluate within 10 minutes, while considering the relevance criteria provided by the given scenario description.

Model	Accuracy (%)	Precision (%)	Recall (%)
<b>Facial Expressions: Documents</b>			
<i>Baseline</i>	50.0	0.0	0.0
<i>SVM Training &amp; Test Set*</i>	57.1	60.8	52.9
<i>SVM (5WC)*</i>	57.6	60.5	56.5
<i>SVM (10WC)</i>	57.9	60.6	57.7
<i>KNN Training &amp; Test Set (Levenshtein, K=15)*</i>	51.7	49.0	51.4
<b>Facial Expressions: Videos</b>			
<i>SVM Training &amp; Test Set*</i>	63.9	63.7	67.0
<i>SVM (5WC)*</i>	64.9	64.5	68.8
<i>SVM (10WC)*</i>	64.4	64.7	69.5
<i>KNN Training &amp; Test Set (Levenshtein, K=15)*</i>	56.9	58.6	51.5
<b>Peripheral Physiological Signals: Documents</b>			
<i>Baseline</i>	50.0	0.0	0.0
<i>SVM Cross-Validation (5 folds)*</i>	60.4	61.9	54.1
<i>KNN Cross-Validation (5 folds, Levenshtein, K=15)*</i>	55.6	49.3	54.3
<b>Peripheral Physiological Signals: Videos</b>			
<i>SVM Cross-Validation (5 folds)*</i>	66.5	66.2	67.6
<i>KNN Cross-Validation (5 folds, Jeffrey Divergence, K=1)*</i>	63.7	65.5	64.4

Table 1: Results for models trained on facial expressions (motion units) and peripheral physiological signals.

### 3. RESULTS

The modeling goal was to predict with reasonable accuracy topical relevance, using sensory data that derive from facial expressions and other peripheral physiological signals as the only feedback information. From the latter signals we extracted a set of features and performed discriminant analysis, using a range of classification methods such as support vector machines (SVM) and nearest neighbour clustering (KNN). We measured the performance of our models using the standard information retrieval metrics of accuracy, precision and recall. Accuracy was computed as the fraction of items in the test set for which the prediction was correct.

For each classification method, we present only the model which achieved the best performance, among the rest in its category. The results are shown in Table 1. The McNemar’s test was applied to check for statistically significant variation against the baseline. Models marked with (\*) got significantly different results, compared to the baseline model, with  $p < 0.005$ . Among all the models, the SVM model held the best performance, giving a reasonable, though rather noisy, prediction of topical relevance. The boosting that we performed, using either 5 or 10 weak classifiers, gave a slight increase in the accuracy. With the exception of one SVM model, all the rest had a statistically significant difference in their classification accuracy, compared to the baseline model. The discriminative KNN model had in all cases the lowest performance.

### 4. CONCLUSIONS

Overall, the evidence that we accumulated validates both of our hypotheses, namely that users’ affective responses, as determined from the observation of their facial expressions and other peripheral physiological signals, will vary across the relevancy of perused information items. This improved the classifiers’ prediction accuracy, compared to the baseline, across the media type. Furthermore, the results of our initial evaluation indicate that prediction of topical relevance is possible, and to a certain extent models can benefit from tak-

ing into account user affective behavior. Our model-based approach was also designed to be as independent as possible from the viewed content and context, therefore, making its application generalizable to a range of different search topics and media types. This is perhaps the most significant contribution of this work, since it will potentially influence other aspects of the retrieval process, such as implicit feedback methods, indexing, ranking and recommendation techniques.

### 5. ACKNOWLEDGMENTS

The research leading to this paper was supported by the European commission, under the contract FP6-033715 (MI-AUCE Project). Our thanks to MKLab for participating in this study.

### 6. REFERENCES

- [1] I. Arapakis, J. M. Jose, and P. D. Gray. Affective feedback: an investigation into the role of emotions in the information seeking process. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 395–402, New York, NY, USA, 2008. ACM.
- [2] M. Lynne, R. C. Sheldrick, and R. Paulette. *Affective dimensions of information seeking in the context of reading*. Medford, NJ: Information Today, 2007.
- [3] H.-R. Pfister and G. Böhm. The multiplicity of emotions: A framework of emotional functions in decision making. *Judgment and Decision Making*, 3:5–17, 2008.
- [4] T. Saracevic. Relevance reconsidered. information science: Integration in perspectives. *Proceedings of the Second Conference on Conceptions of Library and Information Science, Copenhagen, Denmark*, pages 201–218, 1996.
- [5] R. Valenti, N. Sebe, and T. Gevers. Facial expression recognition: A fully integrated approach. *Image Analysis and Processing Workshops, 2007. ICIAPW 2007. 14th International Conference on*, pages 125–130, Sept. 2007.
- [6] S. Vrochidis, P. King, L. Makris, A. Moutzidou, V. Mezaris, and I. Kompatsiaris. Mklab interactive video retrieval system. In *CIVR '08: Proceedings of the 2008 international conference on Content-based image and video retrieval*, pages 563–564, New York, NY, USA, 2008. ACM.