

# The MediaEval 2016 Context of Experience Task: Recommending Videos Suiting a Watching Situation

Michael Riegler<sup>1</sup>, Concetto Spampinato<sup>2</sup>, Martha Larson<sup>3</sup>, Pål Halvorsen<sup>1</sup>, Carsten Griwodz<sup>1</sup>

<sup>1</sup>Simula Research Laboratory and University of Oslo, Norway

<sup>2</sup>University of Catania, Italy

<sup>3</sup>Delft University of Technology and Radboud Univeristy, Netherlands

{michael, paalh, griff}@simula.no, cspampin@dieei.unict.it, m.a.larson@tudelft.nl

## ABSTRACT

In this paper we present an overview of the Context of Experience Task: recommending videos suiting a watching situation which is part of the MediaEval 2016 Benchmark. The aim of the task is to explore multimedia content that is watched under a certain situation. The scope of the this years task lies on movies watched during a flight. We hypothesize that users will have different preferences for movies that are watched during a flight compared to when a movie is watched at home or the cinema. This is most probably influenced by the context and the devices used to watch. In the case of being on a flight, the context is clearly different to normal situation (noise, compact, bad air) and also the devices differ (small screens, bad audio quality). The main goal of the task is to estimate if a person would like to watch a certain movie on the airplane or not. As dataset we provide a large collection of movies, collected from an airline, including pre-extracted visual, text and audio features.

## 1. INTRODUCTION

The Context of Experience Task of MediaEval 2016 tackles the challenge of predicting the multimedia content that users find most fitting to watch in specific viewing situations. A majority of the existing work on video recommendation focus on predicting personal preferences. As such, it overlooks cases in which context might have a strong impact on preference relatively independently of the personal tastes of specific viewers. Particularly strong influence of context can be expected in unusual, potentially psychologically or physically straining situations.

The use case for this task is the case of viewers watching movies on an airplane. Here, independently of personal preferences, viewers share the common goal, which we consider to be a *viewing intent*, of relaxing, passing time and keeping themselves occupied while being confined in the small and often crowded space of an airplane cabin.

Figure 1 gives an impression of a screen commonly used on an airplane and some situations that can occur during a flight that can influence the watching experience of the viewers. Subfigure 1(a) shows the optimal situation without a distraction and a acceptable video quality. Subfigure 1(b) is the same movie but the line of sight is blocked by a stewardess. Subfigure 1(c) depicts the case of when a movie gets stopped during an announcement. For many users this can

be very annoying, especially when it happens for by them as unimportant judged announcements (for example security introduction if already known from before). The final subfigure 1(d) shows the screen with a strong glare effect which is common for airplane screens. Taking this few example into consideration it is clear that a persons experience of a movie during the flight is heavily influenced by the context. This might lead to a different consumption behavior of movies than usual, for example, people might not want to watch a very good movie because they want to experience it in a very good quality and a more pleasant situation at home or the cinema.

The objective of the task is to predict which videos are preferred by the users to achieve the goal of passing time, relax or get distracted. This also includes limitations of the technology being used (e.g., screen size) and the environment itself (e.g., background noise, interruptions, presence of strangers). The use case of the airplane scenario is a situation that is well known by many people and it is also common to use the multimedia services provided by the airlines to make the journey more pleasant and passing faster<sup>1</sup>. It is important to point out that even if the use case for the task is limited to the watching movies on the airplane scenario, the influence of Context of Experience is much broader and also important in other scenarios. Other scenarios of stressful contexts where videos are becoming increasingly important include hospital waiting rooms, public transport, dentists offices and all other kind of situation where people use multimedia to distract them.

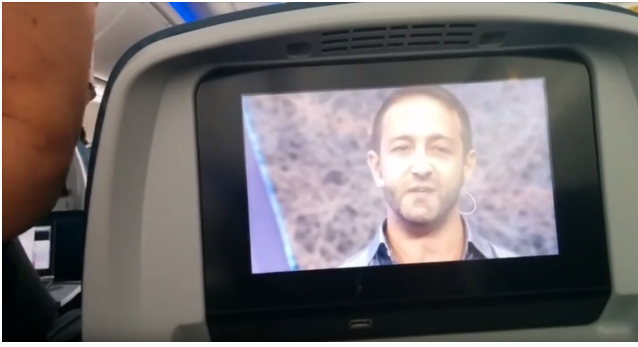
## 2. TASK DESCRIPTION

For the task we provide a dataset containing a list of movies, including links to descriptions and video trailers, pre-extracted features and metadata. The main objective of the task is to classify each movie into *+goodonairplane* or *-goodonairplane* classes. The ground truth for the task has been collected from two sources. The first source is a list of movies used by a major airline<sup>2</sup>. The second source are user judgments on movies that are collected via a crowdsourcing tool<sup>3</sup>. Task participants are asked to form their own hypothesis about what they think is important for users viewing movies on an airplane, and design an approach using appropriate features and a classifier, or decision function.

<sup>1</sup><http://www.tripinsurance.com/tips/guide-to-the-best-moviestv-shows-to-watch-on-a-plane>

<sup>2</sup>[http://www.klm.com/travel/no\\_en/prepare\\_for\\_travel/on\\_board/entertainment/onboard\\_movies.htm](http://www.klm.com/travel/no_en/prepare_for_travel/on_board/entertainment/onboard_movies.htm)

<sup>3</sup><https://microworkers.com/>



(a) The *normal* situation while watching a movie on a plane.



(b) The user gets distracted from the movie by a stewardess serving something to the neighbour passenger.



(c) The movie is stopped for an announcement.



(d) Glare on the screen makes it almost impossible to see what is going on.

**Figure 1:** The four images show the *normal* or *optimal* situation compared to three disturbing situations that can occur during a flight.

The value of Context of Experience lies in the intersection of research efforts currently ongoing in two different disciplines. First, in the field of multimedia, it is related to work on the impact of video content on viewers. Several datasets and benchmarks have contributed to supporting research that develops algorithms capable of automatically predicting the emotional impact (affective impact) of video content on the viewer.

Within the MediaEval benchmark<sup>4</sup>, these have been an early task on predicting viewer experienced boredom [14] and a current task on the affective impact of movies [13]. Moreover, in the field of multimedia, extensive work has been carried out on Quality of Experience, including [6, 5, 3, 4, 2] which is also related to this task. Furthermore, the Context of Experience task is correlated with multimedia research in the area of viewer intent, because the intent of users (i.e., the reason why they want to watch movies on the airplane) is a strong influencing force on what they watch [7]. Finally, in the field of recommender systems, Context of Experience is related to work on context-aware recommendation [1, 12]. There exists a major difference between the challenge of Context of Experience and the challenge of context-aware movie recommendations, that have been presented by researchers before [10, 11]. Context of Experience assumes that the experience of viewing a movie interacts with the context in which a movie is viewed. This leads to the situation that the movie is actually able to *change* the context. Addressing the Context of Experience challenge means not ‘just’ matching movies with personal tastes like in a recommender system, but rather helping users accomplishing goals that they want to achieve by consuming movies.

To tackle the task, it can be addressed by leveraging techniques from multiple multimedia-related disciplines, including social computing (intent), machine learning (classification), multimedia content analysis, multimodal fusion, and crowdsourcing. Further, we hope that it will be useful for content provider, since the exploitation of intent in combination with users’ satisfaction could lead to more sophisticated ways to provide methods of providing a better service to the users.

### 3. DATASET

The dataset is described in more detail in the papers [8] and [9] and it has also been used in the *Multimedia Challenges Beyond Visual Analysis* contest session at the international conference on pattern recognition 2016<sup>5</sup>.

The provided dataset includes movie titles and links, that allow participants to gather online metadata and trailers. We do not provide the video files because of copyright restrictions. Examples are collected based on movie lists from a major international airline, i.e., KLM Royal Dutch Airlines. The final list of movies is a merged set of movies collected between February and April 2015. The video dataset contains both positive and negative samples, carefully sampled in order to create fair and representative positive and negative class. The data is also split into a trainingset and a testset. In order to collect user judgments, we used an existing system that has been built for the purpose of collecting user feedback of this sort. We evaluate systems both, with

<sup>4</sup><http://www.multimediaeval.org/>

<sup>5</sup>[http://gesture.chalearn.org/icpr2016\\_contest](http://gesture.chalearn.org/icpr2016_contest)

respect to the airline’s choice of movies, and the crowd’s choice of airline-suitable movies. Votes about the labels collected by crowdsourcing are considered as the authoritative labels. Therefore, crowdworkers are asked to rank a subset of movies regarding if they would like to watch this video on an airplane or not. This ranking combined to create the class for each movie in the training and test data.

### 3.1 Technical Details

Overall, the data set contains 318 movies. The trailers are collected from IMDB and Youtube. The participants are also allowed to collect their own data such as full length movies, more meta data like for example user comments, etc. The goal of the developed systems should be to automatically identify appropriate content in terms of if a movie should be recommended for being watched on an airplane or not. Therefore, the methods should not require manual or crowdsourced input. The data contains extracted visual, audio and text features. Furthermore, we provide metadata collected from IMDB including user comments. The provided visual features are Histogram of Oriented Gradients (HOG) gray, Color Moments, local binary patterns (LBP) and Gray Level Run Length Matrix. The audio descriptors are Mel-Frequency Cepstral Coefficients (MFCC). The development set contains 95 labeled movies. The test data contains 223 movies without labels. The labels will be made public available after the task has been completed. The baseline is created using a simple tree based classifier and has a precision of 0.629 a recall of 0.573 and a F1 score of 0.6.

### 3.2 Evaluation

For the evaluation, we use the metrics precision, recall and weighted F1 score. We chose this metrics instead of error rate because the task is related to recommendation problems and therefore the recall is an interesting and important part of the evaluation. Precision is computed by true positives ( $tp$ ) divided by  $tp$  plus false positives ( $fp$ ), and recall is computed by  $tp$  divided by  $tp$  plus false negatives ( $fn$ ). The F1 score is the harmonic mean between precision and recall. Negative and positive classes in both datasets are balanced. As the described in the section before labels for the task are only given for the development data at the begin. Participants are asked to submit a predicted class for each movie in the test dataset. The metrics then are calculated and provided to the participants.

## 4. SUMMARY

The task itself can be seen as very subjective and eventually hard to solve and evaluate. Due to its novel use case, we hope to motivate a lot of researchers to have a look into the field of user intent and Context of Experience. Understanding users intent and what users need to have the best experience will be a important part of multimedia research in the future. We hope that the Context of Experience Task can help to raise awareness into the topic but also provide an interesting and meaningful use case to researchers already working in the field.

## 5. ACKNOWLEDGMENT

This work is partly funded by the FRINATEK project "EONS" (#231687) and the BIA project *PCIE* (#235530)

funded by the Norwegian Research Council and by the EC FP7 project CrowdRec (#610594).

## 6. REFERENCES

- [1] G. Adomavicius and A. Tuzhilin. Context-aware recommender systems. In *Recommender systems handbook*. Springer, 2011.
- [2] A. Borowiak and U. Reiter. Long duration audiovisual content: Impact of content type and impairment appearance on user quality expectations over time. In *Proc. of QoMEX*. IEEE, 2013.
- [3] P. Lebreton, A. Raake, M. Barkowsky, and P. Le Callet. Evaluating complex scales through subjective ranking. In *Proc. of Quality of Multimedia Experience (QoMEX)*, pages 303–308. IEEE, 2014.
- [4] B. Rainer and C. Timmerer. A quality of experience model for adaptive media playout. In *Proc. of Quality of Multimedia Experience (QoMEX)*, pages 177–182. IEEE, 2014.
- [5] J. A. Redi, Y. Zhu, H. de Ridder, and I. Heynderickx. How passive image viewers became active multimedia users. In *Proc. of Visual Signal Quality Assessment*, pages 31–72. Springer, 2015.
- [6] U. Reiter, K. Brunnström, K. De Moor, M.-C. Larabi, M. Pereira, A. Pinheiro, J. You, and A. Zgank. Factors influencing quality of experience. In *Proc. of Quality of Experience*, pages 55–72. Springer, 2014.
- [7] M. Riegler, L. Calvet, A. Calvet, P. Halvorsen, and C. Griwodz. Exploitation of producer intent in relation to bandwidth and qoe for online video streaming services. In *Proc. of NOSSDAV*. ACM, 2015.
- [8] M. Riegler, M. Larson, C. Spampinato, P. Halvorsen, M. Lux, J. Markussen, K. Pogorelov, C. Griwodz, and H. Stensland. Right inflight?: A dataset for exploring the automatic prediction of movies suitable for a watching situation. In *Proc. of MMSys*. ACM, 2016.
- [9] M. Riegler, M. Larson, C. Spampinato, J. Markussen, P. Halvorsen, and C. Griwodz. Introduction to a task on context of experience: Recommending videos suiting a watching situation. In *Proc. of the MediaEval 2015*. CEUR-WS.org, 2015.
- [10] A. Said, S. Berkovsky, and E. W. De Luca. Putting things in context: Challenge on context-aware movie recommendation. In *Proc. of CAMRa*. ACM, 2010.
- [11] A. Said, S. Berkovsky, and E. W. De Luca. Group recommendation in context. In *Proc. of CAMRa*. ACM, 2011.
- [12] Y. Shi, M. Larson, and A. Hanjalic. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Comput. Surv.*, 2014.
- [13] M. Sjöberg, Y. Baveye, H. Wang, V. L. Quang, B. Ionescu, E. Dellandréa, M. Schedl, C.-H. Demarty, and L. Chen. The MediaEval 2015 Affective Impact of Movies Task. In *Proc. of the MediaEval 2015*, 2015.
- [14] M. Soleymani, M. Larson, T. Pun, and A. Hanjalic. Corpus development for affective video indexing. *Trans. on Multimedia*, 16(4), 2014.