

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

Michael Riegler¹, Concetto Spampinato², Martha Larson³, Pål Halvorsen¹, Carsten Griwodz¹

¹Simula Research Laboratory and University of Oslo, Norway

²University of Catania, Italy

³Delft University of Technology and Radboud Univeristy, Netherlands

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

ABSTRACT

The Context of Experience Task at MediaEval 2016 is devoted to recommending multimedia content suiting a watching situation. Specifically, the task addresses the situation of viewers watching movies on an airplane. The goal of the task is to use trailer-content and textual metadata in order to estimate whether movies are fitting to watch in flight, as judged by the crowd. The context of an airplane often falls short of an ideal movie-watching situation (noise, lack of space, interruptions, stale air, stress from turbulence) and the device can also impact user experience (small screens, glare, poor audio quality). The task explores the notion that some movies are generally better suited to these conditions than others, and that a component of this suitability is independent of viewers' personal preferences.

1. INTRODUCTION

The Context of Experience Task at the Multimedia Evaluation (MediaEval) 2016 Benchmark tackles the challenge of predicting the multimedia content that users find most fitting to watch in specific viewing situations. When researchers in the area of recommender systems or multimedia information retrieval consider the situations in which viewers consume multimedia content, such as movies, they generally assume comfortable watching conditions. This assumption is understandable, since people do frequently enjoy movies in the quiet, privacy and comfort of their own living rooms, together with friends and loved ones, relaxing in arm chairs and on the couch. However, movie watching is certainly not limited to such situations. In fact, people might choose to watch movies exactly *because* they are in an uncomfortable, stressful situation and would benefit from distraction.

Our ultimate goal is to build recommender systems that support people in finding content that helps them through tough times, i.e., moments at which they are under psychological stress or in physical discomfort. We envisage such contexts to include dentist offices and hospitals. However, here, we focus our effort on on a context that does not involve either physical pain or extreme psychological distress: we chose the context of air travel. Specifically, the Context of Experience Task requires participants to use features derived from video content and from movie metadata in order to predict movies that are appropriate to watch on an airplane.

The next sections of the paper cover related work, and provide more details on in-flight-distractors influencing viewer experience. We close with a brief description of the data set and the task. The description is brief since this information has been provided in detail elsewhere. Specifically, the first description was published in a short paper in the proceedings of MediaEval 2015 [7], which served to launch the task. Additional information was published in [6]. Finally, in order to stimulate cross-benchmark collaboration, the task was also offered as part of the Joint Contest on Multimedia Challenges Beyond Visual Analysis at ICPR 2016, and a paper published that contains a short description and some insights on results [4].

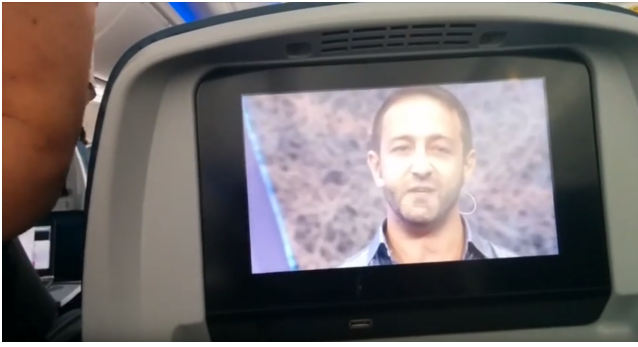
2. RELATED WORK

Although our ultimate aim is to provide viewers with multimedia content for a particular context, we differ from context-aware movie recommendations as addressed by [8, 9]. Context of Experience assumes that the experience of viewing a movie interacts with the context in which a movie is viewed. Instead, we admit that a movie is actually able to *change* the viewer's perception of the context. We emphasize that addressing the challenge of recommending for users' Contexts of Experience means not 'just' matching movies with users' personal taste, but rather also helping users accomplish goals that they want to achieve by consuming movies. These goals may include distracting themselves from discomfort and making time pass more quickly. We also note that the focus of recommender system research on personalization often leads to neglect of cases in which context might have a strong impact on preference relatively independently of the personal tastes of specific viewers, an idea echoed in [5]. Particularly strong influence of context can be expected in the stressful situations that are the focus of our interest.

Context of Experience is obviously closely linked to the area of Quality of Experience of multimedia content. In [10], *Physical context*, *Social Cultural Context* and *Task* are all identified as context-related factors that contribute to the user's perception of quality of experience.

Within the MediaEval benchmark¹, the Context of Experience Task follows upon other tasks that have been devoted to predicting the impact of content on viewers or listeners. These include an Affect Task on predicting viewer experienced boredom [12], the Emotion in Music task [1], a current task on the affective impact of movies [11, 2], and a current task on Predicting Media Interestingness [3].

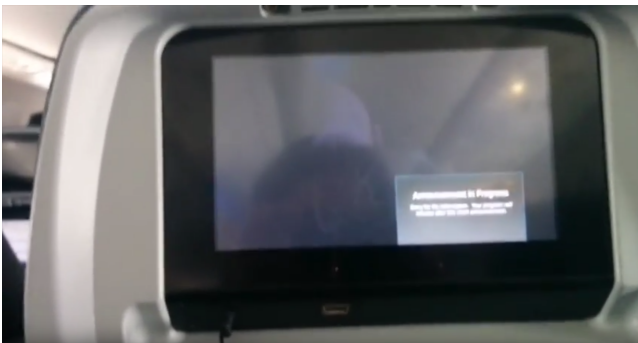
¹<http://www.multimediaeval.org/>



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



(b) A flight attendant serving the neighboring 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 *ideal* situation compared to three distracting situations that can occur during a flight.

3. MOVIES ON A PLANE

On a plane, we assume that movie 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 provides 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. The other subfigures illustrate distracters that impact the movie viewing experience. These examples illustrate how a person's experience of a movie during the flight can be heavily influenced by the context.

4. TASK AND DATA

The objective of the task is to classify each movie as either *+goodonairplane* or *-goodonairplane*. Task participants are asked to form their own hypothesis about what they think is important for people viewing movies on an airplane, and then to design an approach using appropriate features and a classifier or decision function.

The task data set consists of a list of movies, including links to descriptions and video trailers, pre-extracted features and metadata. Movies were collected between February and April 2015 from movie lists of a major international airline, i.e., KLM Royal Dutch Airlines. The set contains an equal number of non-airline movies, sampled with similar distributional properties (e.g., year). We do not provide video files for the trailers because of copyright restrictions. The pre-extracted 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 (MFCCs). Task participants are also allowed to collect their own data such as full length movies, and more metadata, e.g., user comments. The development set contains 95 and the test set contains 223 movies. The data set is balanced 50/50 between *+goodonairplane*/*-goodonairplane*. The ground truth consists of user judgments gathered on CrowdFlower. In total, 548 different workers participated and at least five judgments per movie were collected.

For the evaluation, we use the metrics precision, recall and weighted F1 score. We chose these metrics instead of error rate because the task is related to recommendation. For the purposes of recommendation, a ranked list is often needed. Also, recall is an interesting and important part of the evaluation. A baseline was created using a simple tree based classifier (precision 0.629; recall of 0.573; F1 score 0.6). As mentioned above, more information is available in the other papers that have been published discussing the data set and the task [7, 6, 4]. We hope that the Context of Experience Task can help to raise awareness of the topic and also provide an interesting and meaningful use case to inspire more work in this area.

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] A. Aljanaki, Y.-H. Yang, and M. Soleymani. Emotion in Music Task: Lessons learned. In *this proceedings*.
- [2] E. Dellandréa, L. Chen, Y. Baveye, M. Sjöberg, and C. Chamaret. The MediaEval 2016 Emotional Impact of Movies Task. In *this proceedings*.
- [3] C.-H. Demarty, M. Sjöberg, B. Ionescu, T.-T. Do, H. Wang, N. Q. K. Duong, and F. Lefèbvre. MediaEval 2016 Predicting Media Interestingness Task. In *this proceedings*.
- [4] H. J. Escalante, V. Ponce-López, J. Wan, M. A. Riegler, B. Chen, A. Clapés, S. Escalera, I. Guyon, X. Baró, P. Halvorsen, H. Müller, and M. Larson. ChaLearn joint contest on multimedia challenges beyond visual analysis: An overview. In *Proceedings of the International Conference on Pattern Recognition*, 2016, to appear.
- [5] R. Pagano, P. Cremonesi, M. Larson, B. Hidasi, D. Tikk, A. Karatzoglou, and M. Quadrana. The contextual turn: From context-aware to context-driven recommender systems. In *Proceedings of the 10th ACM Conference on Recommender Systems*, RecSys '16, pages 249–252, 2016.
- [6] 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 Multimedia Systems*, MMSys '16, pages 45:1–45:6, 2016.
- [7] 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 *Proceedings of the MediaEval 2015 Workshop, Wurzen, Germany, September 14-15, 2015*. ceur-ws.org/Vol-1436/Paper5.pdf.
- [8] A. Said, S. Berkovsky, and E. W. De Luca. Putting things in context: Challenge on context-aware movie recommendation. In *Proceedings of the Workshop on Context-Aware Movie Recommendation*, CAMRa '10, pages 2–6, 2010.
- [9] A. Said, S. Berkovsky, and E. W. De Luca. Group recommendation in context. In *Proceedings of the 2nd Challenge on Context-Aware Movie Recommendation*, CAMRa '11, pages 2–4, 2011.
- [10] R. Schatz, T. Hoßfeld, L. Janowski, and S. Egger. *From Packets to People: Quality of Experience as a New Measurement Challenge*, pages 219–263. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [11] 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 *Proceedings of the MediaEval 2015 Workshop Wurzen, Germany, September 14-15, 2015*. <http://ceur-ws.org/Vol-1436/Paper1.pdf>.
- [12] M. Soleymani, M. Larson, T. Pun, and A. Hanjalic. Corpus development for affective video indexing. *IEEE Transactions on Multimedia*, 16(4):1075–1089, 2014.