Simula @ MediaEval 2016 Context of Experience Task

Konstantin Pogorelov, Michael Riegler, Pål Halvorsen, Carsten Griwodz
Simula Research Laboratory and University of Oslo
konstantin, michael, paalh, griff@simula.no

ABSTRACT

This paper presents our approach for the Context of Multimedia Experience Task of the MediaEval 2016 Benchmark. We present different analyses of the given data using different subsets of data sources and combinations of it. Our approach gives a baseline evaluation indicating that metadata approaches work well but that also visual features can provide useful information for the given problem to solve.

1. INTRODUCTION

In this paper we present our solutions for the Context of Experience Task: recommending videos suiting a watching situation [10], which is part of the MediaEval 2016 Benchmark. The Context of Experience task's main purpose is to explore multimedia content that is watched under a certain situation. This situation can be seen as the context under that the multimedia content is consumed. The use case for the task is watching movies during a flight.

The hypothesis is that watching movies during a specific context situation will change the preferences of the viewers. This is related to similar hypotheses in the field of recommender systems as presented in for example [12, 13] where context is also an important influencing factor. Nevertheless, it is also closely related to the field of quality of experience [9, 8, 4] because the context during a flight, such as loud noises and other distractions, can play an important role for which movies viewers chose to watch.

Participants of the context of experience task are asked to classify a list of movies into the two classes, namely, *+goodonairplane* or *-goodonairplane*. To tackle this problem we propose three different approaches. All three methods use information extracted directly from the movies or the metadata containing information about the movies in combination with a machine-learning-based classifier. The remainder of the paper is organized as following. At first, we will give a detailed explanation of our three approaches and the classification algorithm that we used. This is followed by a description of the experimental setup and the results. Finally, we draw a conclusion.

2. APPROACHES

In this section we will describe our three proposed runs in more detail. For all runs we use the same classification algorithm to get the final class.

The classification algorithm that we used for all three runs is the PART algorithm [2], which is based on PARTial decision Trees. PART relies on decision lists and uses a separate-and-conquer approach to create them. In each iteration PART creates a partial decision tree. For each iteration the algorithm finds the best leaf in the tree and uses it as a rule. This is repeated until a best set of rules is found for the given data. The advantage of PART is that it is very simple. The simplicity is achieved by using rule based learning and decision finding that does not require global optimization. A possible disadvantage of the algorithm is that the rule sets are rather big compared to other decision based algorithms such as C4.5 [7] or RIPPER [1].

Nevertheless, for our use case this is not important because the dataset is rather small [11]. For all our runs we use the WEKA machine learning library implementation of PART with the provided (optimal) standard settings [3].

2.1 Metadata

For the metadata only approach we used only metadata provided be the task dataset. We limited the metadata to the following attributes: rating, country, language, year, runtime, Rotten Tomatoes score, IMDB score, Metacritic score, and genre. We pre-processed and transformed rating, language, countries and genre into numeric values for the classification. The different scores for the different movie scoring pages were normalized to a scale from 1.0 to 10. If a value was missing in the dataset we manually searched for the information in the Internet and replaced it with what we found. If we could not find ratings for all scoring services we used 5.0 (average score) as value.

2.2 Visual Information

For the visual data we downloaded the trailers from the provided links and extracted all frames. From each frame we extracted different visual features and combined them into one feature vector for the classification (with a dimension of $3,866$ values).

For the visual features, we decided to use several different global features. The features that we used for this work are: joint histogram, JPEG coefficient histogram, Tamura, fuzzy opponent histogram, simple color histogram, fuzzy color histogram, rotation invariant local binary pattern, fuzzy color and texture histogram, local binary patterns and opponent histogram, PHOG, rank and opponent histogram, color layout, CEDD, Gabor, opponent histogram, edge histogram, scalable color and JCD. All the features have been extracted

MediaEval 2016 Workshop, October 20-21, 2016, Hilversum, Netherlands

Table 1: The configuration of our three submitted runs for the task. R1 combines visual and metadata, R2 uses only the metadata and R3 uses only the visual data for classification. The last row shows the baseline provided by the organizers.

Run	Description
R1	Metadata and visual data combined
R2	Meta data only
R3	Visual data only
Baseline	All available metadata

Table 2: Detailed results for each run and the baseline regarding true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN).

Run	TP	FP	TN	FN
R1	101	65	22	35
R2	127	83	4	9
R3	133	77	10	3
Baseline	78	46	41	58

using the LIRE open source library [5]. A detailed description of all features can be found in [6]

2.3 Metadata and Visual Information Combined

For the final run we combined the metadata with the visual feature information. To combine the visual information with the metadata we first run the classifier on the visual information with a modification so that the output was not binary but a probability for each class. This probability then is added to the metadata as two additional features (probability to be negative or positive). The extended feature vector then is used for finding the final class. This can be seen as a kind of late fusion approach which is in general seen as better performing than early fusion in literature [14].

3. EXPERIMENTAL SETUP

The by the task provided dataset contains all in all 318 movies split into training and testset. For each run we calculated the F1-score, precision and recall. The testset contains 223 movies. For the trailers, only links were provided, and we had to download them. Furthermore, the posters of the movies were also provided but we did not use them in our approaches. Apart form the movies we did also use the provided metadata. We did not collect any additional data such as full length movies, etc and we did not use the pre-extracted visual, text and audio features. The goal of the of the task was, as mentioned before, to automatically identify if a movie is suitable to be watched during a flight or not.

We assessed three different methods executed in three runs. An overview of the conducted runs can be found in table 3 where we provide a summarized overview and short descriptions of each method. The organizers also provided a baseline for comparison based on a simple random tree algorithm (last row in the tables).

Table 3: *MediaEval 2016 Context of Experience Task* official results.

Run	F1-score	Precision	Recall
R1	0.6688	0.6084	0.7426
R2	0.7341	0.6047	0.9338
R3	0.7687	0.6333	0.9779
Baseline	0.6	0.629	0.5735

4. RESULTS

Table 3 gives a detailed overview of the results in terms of true positives, false positives, true negatives and false negatives achieved by our runs and the baseline. Table 3 depicts the official results of the task metrics for our runs and the baseline. All three runs outperformed the baseline significantly. R1 which used metadata and visual information at the same time had the lowest performance. This was surprising for us since we were thinking that this approach would perform best. A reason for the weak performance could be the way of how we combine the different features. The second best of our runs is R2 that uses metadata only. This is not surprising since metadata is well known for performing well and in general better than content based classification. R3 was the best performing approach and even outperformed the metadata approach which was not expected. It seems that for the use case of watching movies on a flight the visual features of the movie play an important role. The reason therefore could be that movies with brighter colors are preferred. Nevertheless, we have to investigate his in more detail to give a final conclusion.

5. CONCLUSION

This paper presented three approaches for the context of experience task, which were able to classify movies into two subsets for being suitable or not to be watched on an airplane. The results and insights gained by evaluating our different methods indicate that there is a difference between what people would like to watch during a flight and that this difference is detectable to a certain extend by automatic analysis of metadata and content based information.

Nevertheless, we would clearly see the need for extending the work by using multiple and larger datasets. Additionally, it might be important to collect user opinions not by crowdsourcing but by actually travelling people.

6. ACKNOWLEDGMENT

This work has been funded by the NFR-funded FRINATEK project "Efficient Execution of Large Workloads on Elastic Heterogeneous Resources" (EONS) (project number 231687) funded by the Norwegian Research Council.

7. REFERENCES

[1] W. W. Cohen. Fast effective rule induction. In *Proceedings of the twelfth international conference on machine learning*, pages 115–123, 1995.

[2] E. Frank and I. H. Witten. Generating accurate rule sets without global optimization. In J. Shavlik, editor, *Fifteenth International Conference on Machine Learning*, pages 144–151. Morgan Kaufmann, 1998.

[3] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[4] P. Lebreton, A. Raake, M. Barkowsky, and P. Le Callet. Evaluating complex scales through subjective ranking. In *Proc. of QoMEX)*. IEEE, 2014.

[5] M. Lux. Lire: Open source image retrieval in java. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 843–846. ACM, 2013.

[6] M. Lux and O. Marques. Visual information retrieval using java and lire. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 5(1):1–112, 2013.

[7] J. R. Quinlan. *C4. 5: programs for machine learning*. Elsevier, 2014.

[8] J. A. Redi, Y. Zhu, H. de Ridder, and I. Heynderickx. How passive image viewers became active multimedia users. In *Visual Signal Quality Assessment*. Springer, 2015.

[9] 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 *Quality of Experience*. Springer, 2014.

[10] M. Riegler, , C. Spampinato, M. Larson, P. Halvorsen, and C. Griwodz. The mediaeval 2016 context of experience task: Recommending videos suiting a watching situation. In *Proceedings of the MediaEval 2016 Workshop*, 2016.

[11] 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.

[12] 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.

[13] A. Said, S. Berkovsky, and E. W. De Luca. Group recommendation in context. In *Proc. of CAMRa*. ACM, 2011.

[14] C. G. Snoek, M. Worring, and A. W. Smeulders. Early versus late fusion in semantic video analysis. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 399–402. ACM, 2005.