TUD-MIR at MediaEval 2011 Genre Tagging Task: Query Expansion from a Limited Number of Labeled Videos

Stevan Rudinac, Martha Larson, Alan Hanjalic

Multimedia Information Retrieval Lab, Delft University of Technology, Delft, The Netherlands

{s.rudinac, m.a.larson, a.hanjalic}@tudelft.nl

ABSTRACT

In this paper we present results of our initial research on genre tagging. We approach the task from information retrieval perspective using a relatively small number of labeled videos in the development set to mine query expansion terms characteristic of each genre. We also investigate which sources of information associated with the videos or extracted from their audio channel, e.g. title, description, tags and automatic speech recognition transcripts yield the highest improvement within our query expansion framework. The experiments performed on MediaEval 2011 Genre Tagging dataset demonstrate the effectiveness of our approach.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *retrieval models*, *query formulation*.

General Terms

Algorithms, Performance, Experimentation.

Keywords

Genre tagging, query expansion, video retrieval.

1. INTRODUCTION

In this paper we present results of our initial research on genre tagging, conducted as part of the participation in MediaEval 2011 benchmark. Aiming to create a solid baseline for the future work, we investigate which sources of information, including automatic speech recognition transcripts and metadata associated with the videos, such as e.g. title, description and tags would yield the best performance in the task. Information about genre is generally encoded in both visual and spoken channel of the video. In the specific case of semi-professional user-generated videos, used to compose MediaEval 2011 Genre Tagging datasets, the visual channel usually doesn't provide enough information to discriminate between videos based on genre [6], because a large number of videos depict a single person talking about a particular topic. For this reason, here we focus on the spoken channel and metadata only, while the possibilities of exploiting visual content to improve performance in a further step are explored in [6].

Motivated by the success of information retrieval approaches to semantic video annotation demonstrated in the Tagging Task Professional and WWW of MediaEval 2010 benchmark [2], we perform genre tagging within an information retrieval framework. We conjecture that, given a relatively small number of videos in the development set, it would be practically infeasible to train a

language model for each individual genre label. Instead, in our approach we take a genre label to be an initial query and mine additional genre-specific query terms from the text associated with the available labeled videos. We choose to use query expansion because in our previous work [4] it has proven effective in semantic-theme-based video tagging and retrieval.

The experiments reported here were performed on MediaEval 2011 Genre Tagging datasets [1], which consist of semi-professional documentary videos downloaded from blip.tv together with the associated metadata. The metadata available with the videos include title, description, tags as well as the id of show to which a particular video episode belongs. The development set consists of 247 videos for which the genre labels are provided. The test set is larger and consists of 1727 videos. Each video in the development and the test set belongs to one of 26 genre categories defined by blip.tv (e.g. *art, autos and vehicles, business, default category* etc.). The task requires prediction of genre label for the videos in the test set. In the following, we first describe our query expansion approaches as well as information sources used. Then, we report on experimental results which confirm effectiveness of our approach to genre tagging and indicate research directions that should be pursued for further performance improvement.

2. APPROACHES

In all official runs, we expand queries using the videos available in the development set. Additionally, we experiment with several other query expansions and report results as "unofficial runs".

2.1 Query Expansion via Labeled Videos

We conjecture that a set of terms characteristic of a particular genre could be extracted even from a small number of labeled videos and further used for query expansion. This concept has been widely exploited in e.g. information retrieval approaches with relevance feedback [5]. In our approach, we treat a genre label as the original query and sample additional query terms from the text associated with the videos of that particular genre available in the development set. For each video, text from information sources used in a particular run is concatenated in a single document and then stopword removal and stemming are applied. Further, for each genre we rank all terms in the development set vocabulary according to the decreasing Offer Weight [3] and extend the initial query (genre label) with the 20 top-ranked terms.

$$OW(i) = r * \log\left(\frac{(r+0.5)*(N-n-R+r+0.5)}{(n-r+0.5)*(R-r+0.5)}\right) \qquad (1)$$

In the formula above, r is the number of videos of a particular genre term $t(i)$ appears in, R is the total number of videos of that genre, N is the total number of videos in the collection and n is the

number of videos in the collection term $t(i)$ appears in. As a retrieval method we use negative divergence between multinomial models of the query and the document (video) implemented in Lemur Toolkit. Since the *default category* is very broad and diverse we produce a ranked list of videos for this genre independently. We conjecture that the videos that are ranked low, or don't appear at all in the results lists produced for the other 25 genres, likely belong to the *default category*. Therefore, we produce the ranked list for this genre according to the increased video score, $VS_i = \sum_{g=1:25} \left(N_g - R_{gi} \right) / N_g$, where N_g is the total number of videos retrieved for a particular genre g and R_{gi} is the rank of video v_i in that list. If a particular video doesn't appear in the results list produced for a genre g, R_{gi} is set to 0.

In the official runs we test how different sources of information influence tagging performance of the approach. Performance of the runs described below is reported in Table 1.

ASR: In the official `run_1` we investigate the scenario when no metadata is available. Expansion terms are sampled from ASR transcripts of videos in the development set and the retrieval is performed on the ASR transcripts of the videos in the test set.

Metadata No Tags: In the official `run_2` only video title and description are exploited.

Metadata & ASR: In the official `run_3` we use title, description, tags and ASR transcripts associated with the videos.

Metadata: To produce results in the official `run_4` we index only title, description and the tags associated with the videos.

Reranking With Show ID: In the case of blip.tv dataset, the show id is a strong genre indicator. Specifically, on the development set we noticed that episodes from the same show are usually of the same genre. However, we decided not to use the show ids in the first 4 official runs because we wouldn't be able to localize performance improvement and isolate contribution of other information used. We use results produced in the official `run_4` as the baseline for reranking, because of the highest performance on the development set. For each genre, we utilize show ids to compute median rank of the videos (episodes) coming from the same show. In the official `run_5`, we follow the general video search reranking idea, ranking videos of the same show together and according to their median rank in the starting results list. This run is meant to complement visual reranking runs from [6] which investigate usefulness of visual channel for discriminating blip.tv videos based on genre. In the unofficial reranking `run_6` we use a similar idea and rank at the top all episodes from the test set belonging to the shows that were in the development set labeled by a given genre label. Videos belonging to the same show, are sorted according to their rank in the initial results list and otherwise alphabetically. The remaining videos from the initial results list are sorted according to their initial ranks. Note that we consider the strength of show id as a genre indicator to be an artifact of this particular dataset and do not expect this approach to generalize

2.2 Baseline Query Expansions

Besides the approach described in the previous section, we run several experiments using unexpanded queries (genre labels – a baseline run) and several query expansions: PRF, WordNet, Google Sets and YouTube. To expand queries via YouTube, we first download metadata (e.g. title, description and tags) of the top-50 ranked videos returned by YouTube for each genre label, except for *default category* and sample 20 expansion terms using the Offer Weight as explained in the previous section. For a description of the baseline retrieval run and the remaining three query expansion approaches please refer to [4]. Such composed queries are used to query metadata associated with the videos in the test set. Videos in *default category* are ranked as described in previous section. To conserve space, we report only the performance of a hypothetical oracle query expansion indicator that chooses the best performing query expansion (PRF, WordNet, Google Sets or YouTube) or the baseline for each genre (unofficial `run_7` in Table 1). Failure analysis confirms that across genres all of these choices make a contribution to performance in individual cases.

Table 1. Performance of reported runs expressed in terms of MAP; officially submitted runs are indicated with "^"

run_1^	run_2^	run_3^	run_4^	run_5^	run_6	run_7
0.2146	0.2699	0.3212	0.3937	**0.4191**	**0.5594**	0.2175

3. DISCUSSION AND CONCLUSIONS

We presented several approaches to genre tagging for web video classification, based on simple and proven information retrieval concepts. The experimental results summarized in Table 1 confirm their effectiveness for the task. We show that it is possible to make effective use of sampling genre-specific expansion terms, even when only a limited set of labeled videos is available. Further, we show that the use of metadata yields the highest performance within our framework. Reranking with show ids (`run_5` and `run_6`) further improves performance, but we have strong reservations about generality of this conclusion, because it might be the artifact of blip.tv portal. It is also interesting to notice that the use of ASR transcripts together with metadata does not improve performance of genre tagging, which is opposite to our earlier findings on "general" video retrieval and tagging [2]. Finally, the results in Table 1 show that the performance of "naïve" baseline, PRF and query expansions using thesauri and collateral corpora is far below level of the approach presented in Section 2.1. In the future we will work on refinement of the approach and investigate the performance on e.g. substantially larger video collections. We will also investigate how visual modality could be exploited for improved genre tagging performance.

4. ACKNOWLEDGMENTS

5. REFERENCES

[1] Larson, M. et al. 2011. Overview of MediaEval 2011 Rich Speech Retrieval Task and Genre Tagging Task. In *MediaEval '11*.

[2] Larson, M. et al. 2011. Automatic tagging and geotagging in video collections and communities. In *Proceedings ACM ICMR '11*.

[3] Robertson, S. E. 1991. On term selection for query expansion. J. Doc. 46, 4.

[4] Rudinac, S., Larson, M. and Hanjalic, A. 2010. Exploiting noisy visual concept detection to improve spoken content based video retrieval. In *Proceedings ACM MM '10*.

[5] Ruthven, I. and Lalmas, M. 2003. A survey on the use of relevance feedback for information access systems. Knowl. Eng. Rev. 18, 2.

[6] Xu, P. et al. 2011. TUD-MM at MediaEval 2011 Genre Tagging Task: Video search reranking for genre tagging. In *MediaEval '11*.