UAB at MediaEval 2011: Genre Tagging Task

Richa Tiwari
The University of Alabama at Birmingham
Birmingham, Alabama, USA
rtiwari@cis.uab.edu

Chengcui Zhang
The University of Alabama at Birmingham
Birmingham, Alabama, USA
zhang@cis.uab.edu

Manuel Montes
National Institute of Astrophysics, Optics and Electronics
Puebla, Mexico
mmontesg@inaoep.mx

ABSTRACT
We describe our approach and results towards the genre tagging task of MediaEval 2011. We approached this as an Information Retrieval task and applied a pseudo relevance feedback (PRF) approach for query expansion. Query expansion was also done using WordNet[1] and Wikipedia[2]. Our results show improvement in the tagging task using these query expansion techniques over the original query retrieval results.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *query formulation, relevance feedback, retrieval model, search process.*

Keywords
Genre video tagging, Information retrieval, query expansion, categorization, pseudo relevance feedback.

1. INTRODUCTION
Video indexing can be improved by tagging the videos with keywords. Genre tags communicate about the content of the videos which can assist users or viewers to decide if the video will interest them or not. In this paper we present our approach and results towards the task of automatically tagging videos with genre labels for MediaEval 2011. Participants were given videos, including speech transcripts, metadata and user tags. The task is to assign one genre tag to each of the videos [2]. This task of genre detection is approached as an Information Retrieval task. There are two main phases in any IR approach; the first is indexing of documents and the second is actual retrieval step. One of the most popular approaches towards indexing the documents is the vector space model [3]. Using the term frequency and inverse document frequency (TF-IDF), we can produce a weight for each term in a document and determine the important terms that can describe it. The second step of retrieval involves query expansion. There are several ways in which query expansion can be approached [4].

WordNet and Wikipedia are often used for query expansion and have shown to produce improvement in retrieval results. Another methodology known as pseudo relevance feedback (PRF) can help in improving the search results by using additional terms from the highly ranked documents for each query, as expansion of those queries [5]. We have used this approach along with expansion using WordNet and Wikipedia in this task.

2. SYSTEM OVERVIEW
We approach the task of automatic genre tagging of the videos in MediaEval 2011 as an IR task in which the videos are considered as the documents to be retrieved, and the genre tags are our queries. We present the results of three official runs in Table 1. All the runs include the same preprocessing of the dataset and the query terms and differ in only the amount of input being used at each run, i.e., only speech transcripts for first run, speech transcripts and metadata for second run and finally also the tags for third run. We also perform some unofficial runs for further result analysis and show their results in Table 3.

2.1 Video Document Indexing
Preprocessing of video documents involves stop word removal and stemming. We use a list of stop words freely available on the web [3]and remove them from the data. Next we use a Porter stemmer[4] to stem the text. We did not do any preprocessing on Non-English text. Once the texts are pre-processed, we calculate TF-IDF for each document and keep the words with non-zero TF-IDF values as the index words for those documents. For Runs 2 and 3 in which we consider metadata containing the title of the video, a similarity bias is given to the video title. For example, there are several episodes of the same show in the dataset and which have similar name. We can assume that different episodes of the same show would most likely belong to the same genre. For example, ***EconomyInCrisis****-CaliforniasGrowingDeficit561* and ***EconomyInCrisis****-FewSignsOfRecovery990* are two different episode videos of the same show (*Economy In Crisis*) with the same starting name and belong to the same genre. Consequently, we believe that different episodes of a show with a genre such as *Politics* may talk about politics of different countries or different political agendas but they would all still have the *Politics* genre. Based on this assumption, if one episode video of a show is ranked at the top for any of the genres by our approach, we give the other episode videos of the same show the same genre.

2.2 Query Expansion
Genre tags (26 unique) are considered as the query terms for this task. We use synonyms from the first synset of the query terms produced by the WordNet, as their expansion. We also retrieve Wikipedia article for each query term and extract useful terms from them. To extract keywords for expanding the query terms using Wikipedia articles, we first compute TF-IDF for each term in each article and then considered the top 40 terms with the highest TF-IDF values as the keywords of those articles. We also expanded the queried using PRF. In this technique the top 50 documents retrieved by using the cosine similarity between query

[1] http://wordnet.princeton.edu

[2] http://www.wikipedia.org

[3] http://www.lextek.com/manuals/onix/stopwords1.html

[4] http://snowball.tartarus.org/algorithms/porter/stemmer.html

MediaEval 2011 Workshop, September 1-2, 2011, Pisa, Italy

documents and the video document vectors are considered as the relevant documents and used for further query term expansion. All the words from these documents are joined with the original expanded query and TF-IDF values for these combined query documents are calculated. These expanded queries are again used to retrieve the remaining video documents using cosine similarity between the query vector and the video document vector. We finally take the top 50 retrieved documents from the second phase, i.e., after PRF, and add them to our earlier list of retrieved document. For Run 2 and Run3 we consider an additional clue based on video titles, i.e., after performing pseudo relevance feedback, we match the show names of the remaining non-retrieved videos (not among the first phase top 50 and also second phase top 50) with the already retrieved videos. If the names match we categorize the new videos in the same genre as the matched videos, given the case that their cosine score is non-zero for that genre. The videos that do not match any genre by doing any of these steps fall in the "default category 1006" genre.

3. EXPERIMENTAL RESULTS ANALYSIS

Table 1 shows the official results of our 3 runs. They differ from each other in that the input video document for Run1 includes only Speech Transcripts, whereas Run2 consists of speech transcripts as well as metadata including the title. Input for Run3 includes everything for Run2 as well as user defined tags for each video.

Table 1. Official results (MAP scores)

	Run1	**Run2**	**Run3**
MAP	0.0621	0.0934	0.094

From Table 1, we can see that the Mean Average Precision (MAP) scores of Run 2 and Run 3 are higher than Run 1 and this can be attributed to a possible reason that we have metadata and extra user assigned tags in them. By looking at the individual query MAP scores, we can see that the queries with maximum MAP scores are *Sports* and *Religion* across all three runs. The genre *Personal_or_Auto-biographical* had 0 MAP score for all the three runs. *Sports and Religion* genres have average number of videos whereas there are only 8 videos in *Personal_or_Auto-biographical* genre. Very few videos in this genre may have an impact on the retrieval result, especially after doing PRF the chances of getting a lot of false negative query terms increases. But, since this is not a classification task, it cannot be interpreted from the above statement that the more videos we have the better chances of their retrieval, because the genre with maximum number of videos, i.e., *Politics* with 552 videos does not have the maximum MAP score. As the main aim in this paper is to study the effect of query expansion in genre detection task, we performed further unofficial tests/runs to analyze the effect of these query expansion techniques used. All of these 6 additional runs use the input for official Run 2. We implement another change in these additional runs as compared to our officially submitted runs, such that we took all the video documents retrieved, i.e., the documents with non-zero cosine scores after the second phase and do not just use the top 50 documents. Table 2 explains the difference between each of these 6 runs. Run 4 is our baseline run in which we do not perform any kind of query expansion and match the video document vectors with single term query documents. The results shown in Table 3 are MAP scores for each of these unofficial runs. By examining the results in Table 3, we can see that there are improvements in MAP scores as we combine various query expansion techniques.

Table 2. Unofficial Run description

Runs	**Description**
Run4	Baseline
Run5	Query expansion using Wordnet
Run6	Query expansion using Wikipedia
Run7	Query expansion using PRF
Run8	Query expansion using WordNet + Wikipedia
Run9	Query expansion with PRF + WordNet + Wikipedia

Table 3. Comparison of various query expansion techniques

	Run 4	Run5	Run6	Run7	Run8	Run9
MAP scores	0.07	0.07 2	0.083 9	0.061 9	0.083 6	0.098 6

In Run7 the MAP score drop as compared to Run4 which can be attributed to the fact that we are using PRF. By analyzing Run7 we saw that only 18% of the videos were correctly identified in the first round whereas, in PRF technique we consider all 100% of them to be correct and used them for further query expansion. As we can interpret this leads to a lot of false positive query terms. There is some improvement in the MAP score (32%), when the query expansion by WordNet and Wikipedia was combined along with PRF (Run9). By expanding the queries using WordNet and Wikipedia, 22% of the videos were correctly identified in the first phase, which aids in increase of MAP score. Hence, we can conclude that PRF alone is not a good technique to use, but combined with other kinds of query expansion, can help in retrieval. However, further analysis is needed to determine if the costs of performing PRF is worth the increase in retrieval results.

4. FUTURE WORK

This work is the beginning for the genre detection task that we would like to undertake in near future. A detail work has to be done to increase the accuracy of this task, such as giving different weights to the different inputs. The most obvious analysis that is needed in this work is inspecting the PRF technique applied such as evaluating the top document cutoff or the combination of documents and query after PRF. Further improvements in the results can be done by using other modalities such as visual along with the textual metadata and speech transcripts. We would also like to approach the genre detection task as a classification task and apply all the modalities to learn different genres.

5. REFERENCES

[1] Larson, M., Eskevich, M., Ordelman, R., Kofler, C., Schmiedeke, S. and Jones, G.J.F. Overview of MediaEval 2011 Rich Speech Retrieval Task and Genre Tagging Task, MediaEval 2011 Workshop, 1-2 September 2011, Pisa, Italy.

[2] Manning, C.D., Raghavan, P., and Schtze, H., 2008. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.

[3] Efthimiadis, N.E., 1996. Query Expansion, *In Annual Review of Information Systems and Technology*, Vol. 31, 1996, 121—187.

[4] Yan, R., Hauptmann, A., and Jin, R., 2003. Multimedia search with pseudo-relevance feedback. *In International Conference on Image and Video Retrieval*, 229–238. Lecture Notes in Computer Science, vol. 2728, Springer.