Category Classification of Educational Videos on YouTube through Machine Learning Approaches using Video Titles

Taewon Yoo
Department of Software Convergence
Kyung Hee University
xodnjs0208@khu.ac.kr

Hyunggu Jung
Department of Software Convergence
Kyung Hee University
hgjung@khu.ac.kr

SUMMARY

- We built and evaluated classification models to classify educational videos on YouTube into subcategories.
- YouTube videos' title was used for model training and testing.
- We revealed that it is possible to classify YouTube educational videos into videos with subcategories.

INTRODUCTION

- Recently, YouTube has become the trending social media platform.
- In the field of education, previous studies have shown that YouTube can improve student learning.
- So, It becomes more important for learners to efficiently identify videos that meet their needs.
- In this study, we created three classifiers to classify the categories of educational videos on YouTube and measured the performance of each classifiers.

METHOD

We focused on building a classification model that classifies "deep learning" videos as "Science & Technology", and "piano" videos as "Music".

1. We collected total 3601 titles from the videos on YouTube using YouTube Data API (see Table 1).

Table 1. Execution Criteria of YouTube Data API

API Execute Parameters	Input Parameters
part	snippet
order	date
q(query)	"deep learning" AND "lecture" "piano" AND "lesson"
relevance language	en
type	video

2. Among those videos, we screened and selected videos at the eligibility stage (see Figure 1).

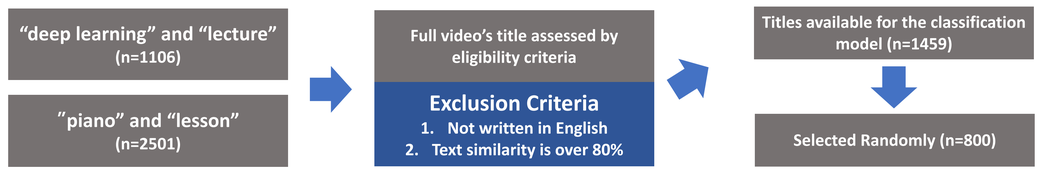

Figure 1. Video Selection Process

3. We preprocessed the text in the title to improve the accuracy of the classification model.

- Removed all emoticons and stop words in the title (e.g., "🎹", "😃" and "💥")
- Performed stemming and lemmatization

4. We used WEKA software as a machine learning tool

- Created Naïve Bayes, SMO-based SVM, C4.5 based decision tree classifier
- Classified videos with stratified 10-fold cross validation using these classifiers

RESULTS

Table 2. Summary of Measured Values of Three Classifiers

	Stratified 10-fold Cross Validation		
	Naïve Bayes	**SVM**	**C4.5**
Correctly Classified Instance (%)	99.875	99.875	99.625
TP Rate	0.999	0.999	0.996
FP Rate	0.001	0.001	0.004

- As shown in Table 2, all three classifiers showed high performance which is over 99% of accuracy.
- The result of this study shows that title of educational videos on YouTube can be used to classify categories of educational videos on YouTube into subcategories.

CONCLUSIONS

- We built classification models to categorize "deep learning" videos as "Science & Technology", and "piano" videos as "Music", and measured models' performances.
- One limitation of this study is that we did not support learners when they identify educational videos.
- Future work will still remains to evaluate learners' satisfaction when the interact with classified videos for educational videos.

ACKNOWLEDGEMENT
This research was supported by the Korean MSIT (Ministry of Science and ICT), under the National Program for Excellence in SW (2017-0-00093), supervised by the IITP (Institute for Information & communications Technology Planning&Evaluation).