

Clip Recommendation based on Topic Similarity

Wonjoo PARK*, Jeong-Woo Son*, Sang-Yun Lee*, Sun-Joong Kim*

*Smart Media Research Group, Broadcasting & Media Research Laboratory, ETRI, Korea

[wjpark, jwson, syllee, kimsj]@etri.re.kr

Abstract— We propose a clip recommendation technology based on topic similarity. Topics of a clip can represent semantics of each contents. When the topic distributions for clips are similar, it means they are alike. In this paper, we propose a system to learn topic distributions for broadcast contents and link clips based on topics similarity of each clip. The higher the similarity is among the clips, the higher the semantic is among them. This system can be adopted clip recommendation with audiences viewing history and their interest.

Keywords— Clip recommendation, topic model, topic similarity, clip metadata, clip search

I. INTRODUCTION

With the advent of the mobile era, there have been many changes in the production, distribution, and consumption ecosystems of broadcast contents. Most of audiences no longer consume media in fixed location and consumes only a part of video rather than the whole broadcast contents. A part of the video is called a clip. It is natural in mobile environment. They do not have much time to watch the whole contents and it requires a large amount of mobile data. The continuing strength of this mobile platform and the content segmentation phenomenon are also affecting service area and research area which share and connect to broadcast contents[1]. Naver and Kakao are providing VoD clip service. In order to provide this service, content providers manually divide the content into clips and tag metadata with the tool.

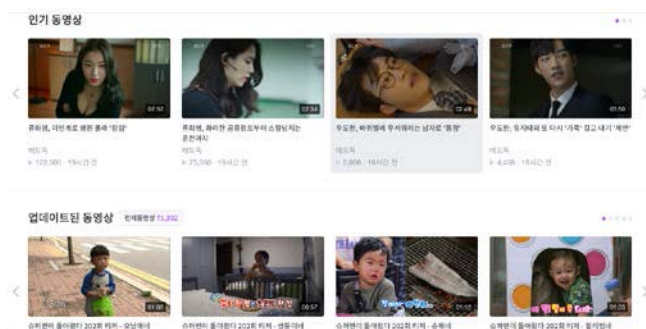


Figure 1. An examples of VoD clip service

Until now, many researches for broadcast contents or movies had focused on video search or recommendation from a large amount of video. They have analyzed subject of content such as synopsis, genre, director, actor, year of production, etc. They have achieved good results in searching and recommending videos. However, according to the change of the media consumption behavior, it is needed to recommend clips according user interests or behaviours. Topic model is a technology to discover semantics of a document by analysing

unstructured and a vast amount of text. It has been used in text mining and language model widely. H. Jeon et al. propose language model based on topic model of LDA for automatic transcription tasks[2]. It is also possible to automatically generate and expand metadata for broadcast contents with topic model. A study on the metadata generation for broadcasting contents uses the text data related to the contents to generate topics[3]. However, because this study only adopted metadata generation for whole contents, there is a limit to use clip search and recommendation. In this paper, to overcome this limit, we propose a system for clip recommendation based on topic similarity. We learn closed caption for a whole broadcast content and segments of them, clips and generate topic distributions for them. Topics for the clip generated by closed caption can represent semantics of each clip. Therefore, when the topic distributions among clips are similar, it is possible to infer that semantic of the clips is alike. We also describe a service that generates topics for real broadcasting contents and analyses their similarities to recommend clips.

II. CLIP RECOMMENDATION BASED ON TOPIC SIMILARITY

In order to reflect trend of clip consumption, we automatically divide broadcast contents into semantic units and call them 'scenes'. Scenes are almost same clips. To generate scenes, we first make shots that are very short-term videos. And then, they are rearranged with the similarity of video, audio, and text features. Also, it must be considered start and end time of each shot. We finally generate scenes with continuous shots[4]. In particular, each scene can have closed caption that is included in a scene. It is used to learn topic distribution for the scene later.

To learn topic distribution for contents, it needs text data associated with them. Scripts or closed captions will be used directly and news and blogs will be used indirectly[5]. The script is a very best document to acquire semantics of broadcast contents. But it has limits. It does not synchronize with the broadcast content and there is no standardized format of the script. These limits require much pre-processing effort. On the other hand, closed captions have synchronized text data with the video, and most of broadcast stations make closed captions almost in real time. But, the closed caption has a limit, too. It has no sufficient text data such time, place, situation, actor, and so on. We select closed caption that is synchronized with the video, because we need start and end time information of scenes. In this paper, the system learns closed caption of an episode unit and generates topic model for the episode. It learns topic distributions of scene with topics for the episode and scenes delivered from automatic scene generator. The system uses a nonparametric model, HDP (Hierarchical Dirichlet Process) to learn topic model for each episode and scenes. The HDP is

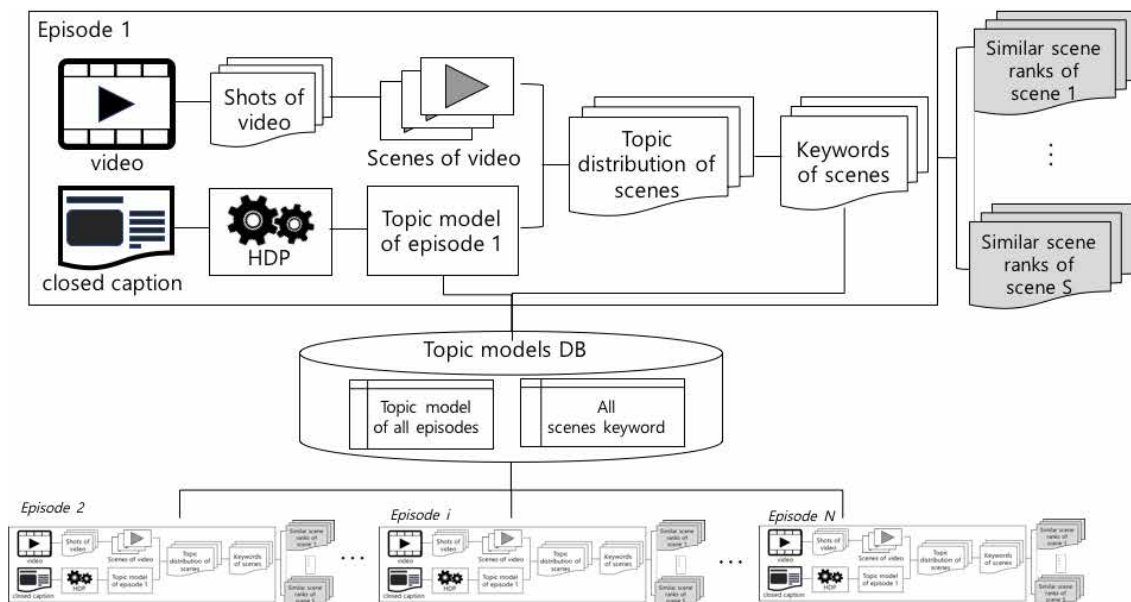


Figure 2. System architecture

approach to clustering grouped data. It uses a Dirichlet process for each group of data, with the Dirichlet processes for all groups sharing a base distribution which is itself drawn from a Dirichlet process[6]. In this system, closed caption is clustering grouped data. Each group of data is caption in each scene. With HDP, we learn topic distributions for episode and scenes.

Figure 2 shows the system architecture to analyze similarity among all scenes. The system generates topic model for episode and scenes. After learning topic model for them, it extracts keywords of scenes. Keywords are extracted with topic

probability. By using inverted index of scene keywords, the similarity is calculated only if the two scenes have one or more same keyword.

We experiment with TV show of TVN, A mysterious and miscellaneous dictionary. For example, 1st episode has 40 scenes and generates 39 topics. Figure 3 is topic distributions of 5th scene in episode 1. The scene has 5 topics and probabilities. Figure 4 is a part of the scene keywords. It has 35 scene keywords, weight and NER. Scene topics and scene keywords can represent semantics of the scene. After the system store scene topics and keywords, the system calculates cosine similarity with topic distributions of scene, scene keywords and keyword's weight among all scenes.

$$\text{Cosine Similarity}(\vec{S_a}, \vec{S_b}) = \frac{\vec{S_a} \times \vec{S_b}}{\|\vec{S_a}\| \times \|\vec{S_b}\|}$$

Sa and Sb are probabilities of scene topics and weights of scene keywords between scene A and scene B. Table 1 shows a similar rank for 5th scene in episode 1. #s is episode ID/scene ID. According to the experimental results, most similar scenes are within the same episode. However, it confirms that the scenes of other episodes are ranked at the top. The top ranked scene contains that people are talking about breakfast, food ingredients, food history etc.

TABLE 1. TOP 10 SIMILAR SCENES OF 5TH SCENE IN EPISODE 1

s#	Top 10 rank									
	1	2	3	4	5	6	7	8	9	10
s#	1/5	1/17	1/24	1/26	<u>5/17</u>	<u>9/31</u>	1/1	<u>2/7</u>	1/32	<u>2/13</u>

```

scene_ID : TVN_T2017-06-00001-05
▼ scene_topic [5]
  ▼ 0 {3}
    topic_ID : 8
    weight : 0.08571428571428572
    topic_labels : 군가 들어오다
  ▼ 1 {3}
    topic_ID : 13
    weight : 0.2857142857142857
    topic_labels : 부르다 원구
  ▼ 2 {3}
    topic_ID : 37
    weight : 0.22857142857142856
    topic_labels : 추적 호주
  ▼ 3 {3}
    topic_ID : 40
    weight : 0.14285714285714285
    topic_labels : 유시민 유희열
  ▼ 4 {3}
    topic_ID : 60
    weight : 0.2571428571428571
    topic_labels : 시키다 놀아나다

```

Figure 3. An examples of topic distribution for a scene

```

scene_ID : TVN_T2017-06-00001-05
▼ scene_keyword_list [35]
  ▼ 0 {3}
    word : 유희열
    weight : 0.049019607843137254
    entity_name : PERSON
  ▼ 1 {3}
    word : 유시민
    weight : 0.0392156862745098
    entity_name : PERSON
  ▼ 2 {3}
    word : 김이
    weight : 0.0392156862745098
    entity_name : NONE
  ▼ 3 {3}
    word : 민
    weight : 0.029411764705882353
    entity_name : NONE
  ▼ 4 {3}
    word : 시키다
    weight : 0.029411764705882353
    entity_name : NONE

```

Figure 4. A part of scene keywords

distribution of scene, word per topic, probability of word[7]. Scene keywords are stored in files and DB. When new scene keywords are extracted, scene keywords are calculated similarity with old scenes' keywords value and keywords

KBS has produced pilot service, clip recommendation based on topic similarity. Figure 5 shows the screen of the service, TVUT. In this service, KBS had served clip recommendation with 'Mysteries of the Human Body' by using a second screen



Figure 5. KBS Pilot service for clips recommendation

service. To suggest related scenes to audiences based on topical relevances, our system supplies information like topics belong to a scene, similar topics in other scenes, and rankings of scenes based on topic similarities.

III. CONCLUSIONS

In this paper, we proposed a system for clip recommendation based on topic similarity. We generated topic model for broadcast content and clips with closed caption. With topic distributions of clips, the system linked them based on scenes similarity. This system can be adopted clip recommendation with audiences viewing history and their interest. We also described a service that generates topics of real broadcasting contents and analysed their similarities to recommend clips.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) [2015-0-00219, Development of smart broadcast service platform based on semantic cluster to build an open media ecosystem]

REFERENCES

- [1] Choi Sun-young, Mobilization of broadcast video ecosystem: share, connect, participate, Broadcasting Trend and Insights, Vol.03, 2015.10.
- [2] H. Jeon, and S. Lee, "Language Model Adaptation Based on Topic Probability of Latent Dirichlet Allocation", ETRI Journal, Volume 38, June, 2016, pp. 487-493.
- [3] Noh Yun-seok, Kwak Chang-uk, Kim Sun-joong, Park Seong-bae, Lee Sang-Jo, A Case study of broadcast scenario analysis using topic model, HCLT conference, 228-230, 2015

- [4] Son, Jeong-Woo, Jeon Junkey, Alex Lee, Sun-Joong Kim. Spectral Clustering with Brainstorming Process for Multi-View Data, Thirty-First AAAI Conference on Artificial Intelligence, 2017
- [5] Kwak Chang-uk, Kim Sun-joong, Park Seong-bae, Kim Kwon-yang, Topic Expansion based on infinite vocabulary online LDA topic model using semantic correlation information, 22(9), 461-466, 2016.
- [6] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." Journal of machine Learning research 3.Jan: 993-1022, 2003.
- [7] Qiaozhu Mei, Xuenhua Shen, ChengXiang Zhai, Automatic labeling of multinomial topic models, KDD'07, 490-499, 2007



Wonjoo Park received her MS degrees in information and communication engineering from Chungnam National University, Daejeon, Rep. of Korea in 2000. She joined ETRI, Rep. of Korea in 2000, where she is currently senior researcher. Her research interests include data mining, topic model, and ontology.



Jeong-Woo Son received his MS and Ph.D. degrees in computer engineering from Kyungpook National University, Daegu, Rep. of Korea in 2007 and 2012 respectively. Since 2013, he has been with Broadcasting & Telecommunications Media Research Lab., Electronics and Telecommunications Research Institute (ETRI), Daejeon, Rep. of Korea. He focuses on machine learning, NLP, and information retrieval.



software, smart TV software platform, and Web technology.

Sang-Yun Lee received his MS and Ph.D. degrees in electronics and telecommunications engineering from Hanyang University, Seoul, Rep. of Korea in 1996 and 2007 respectively. Since 2011, he has been with Broadcasting & Telecommunications Media Research Lab., Electronics and Telecommunications Research Institute (ETRI), Daejeon, Rep. of Korea. He focuses on embedded



control, smart TV, content knowledge mining..

Sun-Joong Kim received her BS degree in computational statistics and her MS degree in computer science from Chungnam National University, Daejeon, Rep. of Korea, in 1989 and 2000 respectively. In February 1989, she joined ETRI, Daejeon, Rep. of Korea, where she is currently principal researcher and director. Her research interests include convergence service