Detecting Movie Segments Using Gaussian Mixture Models for VOD Lectures with Japanese Subtitles

Nobuyuki Kobayashi*, Noboru Koyama**, Hiromitsu Shiina** and Fumio Kitagawa**

(Received 4 June 2011 and accepted in revised form 9 September 2011)

Abstract A search system for VOD (video on demand) lectures is useful if it can be applied beyond searches of only text. To facilitate better searching for movie segments to be used in VOD lectures with Japanese subtitles, we propose a method using subtitles and a solving maximum likelihood estimation from a mixture of Gaussian distributions. The detection was performed by a statistical method by using the expectation-maximization algorithm. This allows for the detecting of movie segments and determining their number. In addition to improving evaluation of movie segments, we provide movie segment rankings. Movie segment rankings are computed for each movie segment using a method that removes one Gaussian distribution from a mixture of Gaussian distributions.

Keywords: VOD learning, GMM, Japanese subtitle, movie search

1. Introduction

In recent years, e-learning systems have proliferated on both the Internet and organizational intranets. E-learning systems such as video on demands (VOD) and web-based training systems are used in many universities. In Okayama University of Science, e-learning systems that employ VOD have been used since 2004(1). Many students are attending lectures that are delivered using the VOD service.

Many VOD lectures use learning management systems for delivering and managing student’s education and the systems are currently under intensive research. The sharable content object reference model (SCORM)(2) is a common standard format used for e-learning contents. In addition to delivering and managing education, consideration should be given to supports for students in understanding VOD lectures. Further, it is necessary to develop a self-learning system which can be used after attending lectures. In particular, the following features are needed: (a) summary outlining lectures; (b) topic searching including location by relevance to homework and the learner’s interest; (c) providing other information on the relationship to previous lectures.

As fundamental research for topic searching, we focus on the detection of movie segments from VOD lectures using search words. To achieve this aim, we proposed a method for detecting movie segments by adopting quadratic functions to generate a histogram containing the frequency distribution of a search word in Japanese subtitles(3). However, this method cannot determine the correct number of movie segments. Therefore, we propose a method to determine the correct number of movie segments by adapting a Gaussian mixture model for VOD lectures with Japanese subtitles. A statistical expectation-maximization (EM) algorithm(4–6) can be used to estimate the parameters of the model. This algorithm fits the frequency distribution of a search word with a Gaussian mixture model and detects a movie segment from each Gaussian distribution (normal distribution). In particular, the characteristic of this research is to use the subtitles to estimate the number of movie segments. In addition, we provide movie segment rankings by an evaluation method that removes one normal distribution from a mixture of Gaussian distributions.

2. E-Learning and Search Systems

In Figure 1, we show a screenshot of a VOD system. It displays a movie of a lecturer on the upper left of the screen, the list of contents on the lower left side, and a slide on the right side. Each lecture consists of three sections. The period of one section varies from about 20 to 30 min. Furthermore, homework regarding the contents of the lecture is given at the end of each lecture. Homework provides a way to test the student’s understanding of the lecture. In this research, we have added a
new search system based on the histogram of the search word frequency in each interval, as shown on the right side of Figure 1. This search system also displays the movie segments and their rankings based on the search words.

3. Estimation of Word Frequency Distribution

In this research, we suppose that one simple histogram peak corresponds to one topic of a movie segment and several simple curves synthesize an approximate curve for a histogram. For example, we use the search word “広告 (advertisement)” in the 14th lecture of the database in the cyber campus at Okayama University of Science. The keyword is used in the 14th homework assignment: “In the business model on the Internet, please discuss the advantages of the keyword advertisement. In addition, please think about how to get revenue”. The graph of the word frequency histogram and the approximate curve are shown in Figure 2. The horizontal shows the movie length. The histogram has a bin width of 1 min. Thus, the vertical axis shows the frequency of words once every minute. As shown in the figure, we can consider that the histogram is synthesized by four distinct peaks.

4. Detection of Movie Segments by a Mixture of Gaussian Distributions

Accordingly, we consider that the approximate curve is synthesized by a mixture of Gaussian distributions from the histogram of a search word. The parameter estimation of a Gaussian mixture model is determined using an EM algorithm. Because it is a linear combination of Gaussian distributions (normal distributions), it is possible to estimate a movie segment from the estimated mixture of Gaussian distributions. The definition of the appearance-time of a search word, the Gaussian distribution and the mixture of Gaussian distributions are as follows.

(1) Let \( n \) be the number of search words and \( X = \{x_1, \ldots, x_n\} \) be a set of appearance times.

(2) The Gaussian distribution (normal distribution) is given as:

\[
\phi(x; \mu_j, \sigma_j^2) = \frac{1}{2\pi\sigma_j^2} \exp\left(-\frac{(x-\mu_j)^2}{2\sigma_j^2}\right).
\]

(3) The mixture of Gaussian distributions is then given as:

\[
q(x; \theta_m) = \sum_{j=1}^{m} w_j \cdot \phi(x; \mu_j, \sigma_j^2), \quad \sum_{j=1}^{m} w_j = 1.
\]

The number of component Gaussian distributions of the Gaussian mixture model is shown by \( l = 1, \ldots, m \) and the parameters of the component Gaussian distributions are \( \theta_m = \{\mu_1, \ldots, \mu_m, \sigma_1^2, \ldots, \sigma_m^2\} \), where the weight of the \( l \) th Gaussian distribution is \( w_l \), the average of the \( l \) th Gaussian distribution is \( \mu_l \), and the variance of the \( l \) th Gaussian distribution is \( \sigma_l^2 \).

4.1 Algorithm for the detection of movie segments

4.1.1 EM algorithm

Parameters of the Gaussian mixtures are estimated by the EM algorithm.

(1-1) Initial step: Let \( \mu_l \) be the midpoint which divides movie times in \( l \), \( \sigma_l^2 = 1 \) and \( w_l = 1 \).

(1-2) E-Step

\[
4.1.2 Detection of movie segments

The Gaussian mixture model of the word frequency distribution which comprises the appearance time for the word is approximated by the EM algorithm. In our VOD lecture, because each section of the lecture is anywhere between 20 and 30 min, we assume that there is a maximum of five movie segments.

These are approximated with the superposition of \( m \), \( m = 1, \ldots, 5 \), kinds of Gaussian distributions. From each component’s Gaussian distribution of the Gaussian mixture model, we estimate the movie segment using a standard deviation of \( \sigma_i \) width from the average \( \mu_i \), which covers 68.26%.

4.2 Example of detected movie segments

As an example of detected movie segments, we show an example using the search word “広告” (advertisement). Table 1 shows movie segments detected a mixture of four Gaussian distributions. The best fitting one is the Akaike’s information criterion (AIC) \(^{(6)}\). Table 2 shows the topics and the movie segments of the actual VOD movie. We detect four movie segments (Table 1), but the actual movie consisted of five topics. From Table 2, we conclude that Segments 1 and 2 correspond to Topic numbers 1 and 2, respectively. In addition, Segments 3 and 4 correspond to Topic Numbers 4 and 5, respectively. However, we cannot detect the movie segment corresponding to Topic number 3.

5. Computation of Movie Segment Rankings

In this research, we consider some evaluation methods to rank movie segments. The main point of the distribution, which removes the \( t \) Gaussian distribution from \( m \) kinds of Gaussian distributions, is defined by the following formula.

\[
q^*_k(x; \theta_m) = \sum_{j=1, j \neq k}^m w_j \cdot \phi(x; \mu_j, \sigma_j^2).
\]

5.1 An approximation of the word frequency by kernel density estimation

Initially, we use the kernel density estimation with the Gaussian kernel function. The definitions of kernel
(1) $b$ is the bandwidth. In this research, we use the band-
width $b=0.75$, which is the width distribution in 1.5 min.
(2) Gaussian kernel function
$$K(x) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} x^2 \right).$$
(3) Kernel density estimator
$$KD(x) = \frac{1}{n \cdot b} \sum_{i} K \left( \frac{x-x_i}{b} \right).$$

5.2 Ranking of a movie segment

The evaluation of each movie segment computes the influence of the removed distribution on the entire distribution during the movie time $t=\mu-\sigma, \ldots, \mu+\sigma$. In this research, we propose three evaluations.

EV1: Kullback–Leibler divergence of a mixture of Gaussian distributions

The characteristic of EV1 is the average of the Kullback–Leibler divergence\(^{(6)}\), which is the difference of the two density functions from the mutual function.

$$V_i(x, \theta_m, k) = \frac{1}{2} \left( \sum_{x=\mu-\sigma}^{\mu+\sigma} q(x, \theta_m) \log \frac{q(x, \theta_m)}{q_i(x, \theta_m)} \right) + \sum_{x=\mu-\sigma}^{\mu+\sigma} \left( q_i(x, \theta_m) \log \frac{q_i(x, \theta_m)}{q(x, \theta_m)} \right).$$

EV2: Sum of the squares of differences

Table 3. The Detected Movie Segments for the Search Word “広告” (advertisement).

| Segment numbers | (Start, End) [mm:ss] |
|-----------------|-----------------------|
| 1               | (9:26, 35:54)         |
| 2               | (16:40, 23:17), (22:54, 29:46) |
| 3               | (16:43, 21:55), (18:31, 28:51), (23:40, 30:34) |
| 4               | (17:10, 20:17), (21:15, 21:57), (24:55, 26:56), (24:55, 26:58) |
| 5               | (17:10, 20:17), (21:15, 21:57), (23:41, 30:14), (25:55, 26:58), (28:40, 31:18) |

Table 4. The Movie Segments Computed Using Quadratic Functions\(^{(3)}\).

| Segment numbers | (Start, End) [Unit: Minute] |
|-----------------|----------------------------|
| 2               | (17.00, 33.00)             |
| 3               | (17.00, 24.00), (25.00, 33.00) |
| 4               | (17.00, 24.00), (25.00, 33.00), (32.00, 33.00) |
| 5               | (17.00, 24.00), (25.00, 33.00), (32.00, 33.00), (32.00, 33.00) |

Figure 3. Ranking by EV1.

Figure 4. Ranking by EV2.

Figure 5. Ranking by EV3.
EV2 evaluates the average of two totals squared. The first term is the difference between the kernel density function and the mixture of the Gaussian distributions. The second term is the difference of the distribution which removes the $k$th Gaussian distribution and the mixture of the Gaussian distributions.

EV3: Difference of Kullback-Leibler divergence

EV3 computes the Kullback-Leibler divergence of the movie segments every second.

Table 5. Ranking by Evaluations.

| Number of distributions | EV1 | EV2 | EV3 |
|-------------------------|-----|-----|-----|
| $k$                     | 1   | 2   | 3   |
| $m=2$                   | 1   | 2   | 3   |
| $m=3$                   | 3   | 8   | 11  |
| $m=4$                   | 6   | 9   | 12  |
| $m=5$                   | 7   | 10  | 14  |

Table 6. Difference of between Rankings by Estimations and Human Judgment.

| Search words          | Frequency | Average | Variance |
|-----------------------|-----------|---------|----------|
|                       | EV1       | EV2     | EV3      | EV1       | EV2     | EV3      |
| Access (アクセス)     | 44        | 4.00    | 3.73     | 4.00      | 19.57   | 14.86    | 16.07    |
| Site (サイト)         | 50        | 4.40    | 4.67     | 4.53      | 10.92   | 10.54    | 11.69    |
| Search engine (検索エンジン) | 24        | 3.07    | 3.60     | 3.60      | 10.92   | 10.54    | 11.69    |
| Keywond (キーワード)  | 34        | 3.87    | 4.53     | 4.00      | 7.70    | 13.87    | 11.14    |
| Advertisement (広告) | 35        | 3.87    | 4.93     | 4.27      | 15.98   | 16.21    | 15.35    |
| Total                 | 187       | 19.20   | 21.73    | 20.13     | 70.14   | 71.97    | 71.94    |

EV2 evaluates the average of two totals squared. The first term is the difference between the kernel density function and the mixture of the Gaussian distributions. The second term is the difference of the distribution which removes the $k$th Gaussian distribution and the mixture of the Gaussian distributions.

\[
V_2(x, \theta_n, k) = \frac{1}{2} \sum_{x \in \mathcal{X}} \left( \left( KD(x) - q(x, \theta_n) \right)^2 \right)
+ \sum_{x \in \mathcal{X}} \left( KD(x) - q_k^*(x, \theta_n) \right)^2.
\]

EV3: Difference of Kullback-Leibler divergence

EV3 computes the Kullback-Leibler divergence of the movie segments every second.

\[
V_3(x, \theta_n, k) = \frac{1}{2} \sum_{x \in \mathcal{X}} \left( KD(x) \cdot \log \frac{KD(x)}{q(x, \theta_n)} \right)
+ \sum_{x \in \mathcal{X}} \left( KD(x) \cdot \log \frac{KD(x)}{q_k^*(x, \theta_n)} \right).
\]

5.3 An example of ranking by evaluations

We show an example of experimental evaluation as an estimation of movie segments of the search word “広告 (advertisement)”.

5.3.1 The approximation using the Gaussian mixture model

The appearance time of the search word “広告 (ad-
5.3.2 Detecting movie segments by each Gaussian distribution

The movie segments are estimated by the average and variance of each Gaussian distribution of the Gaussian distributions. The detected movie segments of each use of the word “advertisement” are shown as a function and the mixture of the Gaussian distributions and the total of the square of the difference of the distribution which removes the $k$th Gaussian distribution and the mixture of the Gaussian distributions (Table 3).

5.3.3 Ranking of detected movie segments

Table 4 shows the movie segments obtained using the quadratic functions. The results of the ranks with-

---

### Table 7. User Evaluation of Movie Segments by EV1.

| Question                                                      | Search words             | Ranking of segments (Start, End) | Strongly agree | Agree | Disagree | Strongly disagree |
|---------------------------------------------------------------|--------------------------|-----------------------------------|---------------|-------|----------|------------------|
| Were you satisfied with identified movie segment by based on search word? | Access (アクセス) Frequency=4 | ① (00:14, 00:54)                   | 1             | 1     | 2        | 1                |
|                                                               |                          | ② (31:11, 31:54)                  | 0             | 1     | 1        | 3                |
|                                                               |                          | ③ (00:00, 09:04)                  | 0             | 1     | 1        | 3                |
|                                                               |                          | ④ (00:00, 14:12)                  | 0             | 1     | 0        | 4                |
|                                                               |                          | ⑤ (00:00, 25:28)                  | 0             | 0     | 1        | 4                |
| Were you satisfied with identified movie segment by based on search word? | Site (サイト) Frequency=50 | ① (00:14, 37:52)                   | 2             | 2     | 1        | 0                |
|                                                               |                          | ② (07:04, 16:35)                  | 0             | 5     | 0        | 0                |
|                                                               |                          | ③ (01:30, 04:56)                  | 2             | 2     | 1        | 0                |
|                                                               |                          | ④ (00:00, 13:34)                  | 2             | 2     | 1        | 0                |
|                                                               |                          | ⑤ (02:38, 41:16)                  | 1             | 2     | 2        | 0                |
| Were you satisfied with identified movie segment by based on search word? | Search engine (検索エンジン) Frequency=24 | ① (00:00, 05:13)                   | 0             | 1     | 1        | 3                |
|                                                               |                          | ② (00:00, 10:56)                  | 0             | 3     | 1        | 1                |
|                                                               |                          | ③ (13:23, 20:17)                  | 0             | 3     | 2        | 0                |
|                                                               |                          | ④ (30:35, 30:37)                  | 0             | 2     | 3        | 0                |
|                                                               |                          | ⑤ (30:35, 30:38)                  | 0             | 3     | 2        | 0                |
| Were you satisfied with identified movie segment by based on search word? | Keyword (キーワード) Frequency=34 | ① (07:02, 37:50)                   | 0             | 0     | 0        | 5                |
|                                                               |                          | ② (03:37, 10:58)                  | 1             | 2     | 2        | 0                |
|                                                               |                          | ③ (06:37, 08:02)                  | 2             | 3     | 0        | 0                |
|                                                               |                          | ④ (06:48, 07:42)                  | 0             | 3     | 2        | 0                |
|                                                               |                          | ⑤ (16:20, 35:20)                  | 0             | 0     | 1        | 4                |
| Were you satisfied with identified movie segment by based on search word? | Advertisement (広告) Frequency=35 | ① (17:28, 24:14)                   | 1             | 3     | 0        | 0                |
|                                                               |                          | ② (17:25, 20:13)                  | 2             | 2     | 1        | 0                |
|                                                               |                          | ③ (18:17, 20:17)                  | 2             | 3     | 1        | 0                |
|                                                               |                          | ④ (23:42, 31:50)                  | 1             | 4     | 0        | 0                |
|                                                               |                          | ⑤ (25:36, 27:40)                  | 1             | 2     | 2        | 0                |
out the component $m=1$ are shown in Table 5. The proposed method is effective for detecting movie segments.

The Gaussian distributions and the ranking by each evaluation of movie segments are shown in Figures 3–5. Each rank by the evaluation of movie segments is shown by the circled numbers.

5.4 Experimental results

We computed the difference between the evaluations of five search words and that by human judgment of the 14th lecture from the database. The averages and variances of the results are shown in Table 6 and Figures 3–5. EV2 and EV3 have a similar rank, whereas EV1 has a different trend. EV2 and EV3 use the difference from the sample kernel density. From the experimental results, we consider that EV1 offers the best evaluation and EV2 gives the worst one. In addition, EV1 tends to choose large segments; EV2 and EV3 tend to choose small segments.

6. Evaluation of Movie Segment Detection by Users

We evaluated the ranking by EV1 from the viewpoint of users. As a preliminary evaluation, we examined the top five movie segments obtained with the search words “access” (アクセス), “site” (サイト), “search engine” (サーチエンジン), “keyword” (キーワード) and “advertisement” (広告). The subjects were five students in the graduate school of informatics.

The results of the questionnaires are summarized in Table 7. We observe that the best fit for the detected movie segments tends to have many word frequencies. The frequencies have a significant influence on the evaluation. In the detection of similar movie segments, the frequencies tend to be more divided. These results are in line with our expectations.

7. Conclusions and Future Work

In this research, we proposed a method to detect movie segments from VOD lectures with Japanese subtitles. This method applies an EM algorithm for the search word frequency and makes an approximation using a Gaussian mixture model. Therefore, we provide information for self-learning and homework tips to students after they attend a lecture. Although this research is just a first step in the development of topic searches, we believe that the proposed method provides the learner with effective information about search words. The distribution of a search word and movie segments of the topic may be effective in understanding the topics. As future tasks, we will consider using information from electronic presentations or slides to obtain a more precise detection of movie segments, and we will develop a search system by related words for movie segments. These will support in achieving our goal of providing better support to students for understanding lectures.

Acknowledgements

This research was supported by Japanese Grant-in-Aid for Scientific Research (C) 22500950.

References

(1) Kitagawa, F. and Onishi, S.: “An Experiment on Selective Course with Face-to-Face or/and E-learning, and Students Behavior”, Jpn. Society of Educational Information, Vol. 22, No. 3, pp. 57–66 (2007) (in Japanese).
(2) Fallon, C. and Brown, S.: e-Learning Standards, St. Lucie Press, Boca Raton, FL (2003).
(3) Kobayashi, N., Shinya, H. and Kitagawa, F.: “A Search System by Word Frequency Distribution for a VOD Lecture with Subtitles”, IPSJ SIG Technical Report, Vol. 2009-DBS-149, No. 7, pp. 1–7 (2009) (in Japanese).
(4) Dempster, A. P., Laird, N. M. and Rubin, D. B.: “Maximum Likelihood from Incomplete Data via the EM Algorithm”, J. Royal Statistical Society Series B, Vol. 39, No. 1, pp. 1–38 (1977).
(5) Duda, R. O., Hart, P. E. and Stork, D. G.: Pattern Classification, Wiley–Interscience, New York (2000).
(6) Bishop, C. M.: Pattern Recognition and Machine Learning, Springer, New York (2006).
Nobuyuki Kobayashi received B.E. and M.E. from Okayama University of Science in 1992 and 1999. He is currently a lecturer of Faculty of Human Sciences, Sanyo Gakuen University. His research areas included natural language processing, search algorithm and e-Learning.

Noboru Koyama is master student in Informatics from Okayama University of Science now. His research areas included statistical algorithm and data mining.

Hiromitsu Shiina received B.E., M.E. and Ph.D. from Toyohashi University of Technology in 1992, 1994 and 1998, respectively. He is currently a lecturer of Faculty of Informatics, Okayama University of Science. His research areas included formal language theory, natural language processing and search algorithm.

Fumio Kitagawa is currently Professor in the Faculty of Informatics, The Okayama University of Science. He graduated in 1980 from The Tsukuba University. He received his Ph.D. degree from The Hiroshima University in 1990. His academic career began in 1982 as a research assistant at the information processing center of The Hiroshima University. From 1995, he is working at The Okayama University of Science. His current interests are in Web based Information Systems and human-machine interface with censor technologies.