混合ベータ分布モデルを用いた共起語頻度によるVOD講義の関連映像区間の検出

中村　槙吾[1]　　坂根 耕平[1]　　椎名 広光[2]

[1] 岡山理科大学大学院 総合情報研究科 情報科学専攻

[2] 岡山理科大学 総合情報学部 情報科学科

626653@teamgear.net[1], i13im01sk@std.ous.ac.jp[1], shiina@mis.ous.ac.jp[2]

1　はじめに

現在，インターネット環境を利用して講義を行うVOD講義が多くの大学で行われている [1][2]．しかしながら，現状のシステムではVODの内容に対する検索機能がほとんど作成されていないため，講義のタイトルからいくつか候補を選び，動画を再生して目的のコンテンツを探す必要がある．そこで本研究では，利用者が重要なポイントや復習したいポイントをより容易に探すことができるようにする検索機能を作成することを目標とする．本研究で開発している検索機能は，字幕データに対する検索語の出現頻度をもとにし，検索語が現れる確率に混合正規分布 [6],[7] や混合ベータ分布 [6],[7] に当てはめ，得られる近似分布 [8],[9] の成分から利用者の意図する映像区間の推定結果を提供する．

また，検索語に対して字幕の同一文に現れる単語である共起語についても出現頻度を求め，検索語の出現頻度と同様に混合ベータ分布モデルをあてはめ，元の検索語と共起語の間の共通映像区間や前後関係を調査した結果についても述べる．

2　VOD システムによる e-Learning 講義システム

本研究で作成しているシステムは，岡山理科大学を含む関連6大学で構成している教育コンソーシアムにおける単位互換制度を利用したVODによるe-Learning講義のシステム [1] 上 (図1) に別途追加する形で開発している．

VODの実行画面は図1のような構成で，左上に講師の動画，左下にそのセクションの内容を表示する．画面の右側に講義資料となるスライドを表示する構成

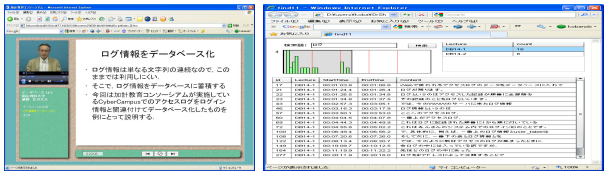

図 1: VOD 講義と検索画面

になっており，ボタンで他のスライドに切り替えることができる．2007年度データベースの講義では，1回の講義は3つのセクションに分かれており，1つのセクションは20〜30分程度となっている．また，各セクションの最後に講義内容に関する課題があり，講義内容の理解を確認するために用いられている．

3　混合正規分布による映像区間推定

字幕データに対する単語頻度から作られるヒストグラムの山の推定に，混合正規分布を使う場合，講義の1セクションに検索語に対する複数の映像区間があると仮定し，混合正規分布によって単語の出現確率を近似する．映像区間の推定はEMアルゴリズムによって推定された混合正規分布によって区間推定を行う．混合正規分布は正規分布の線形結合によって作られるので，正規分布の山を一つの検索語の話題の区間として，このときの正規分布から区間推定を行う．次に検索語の出現時間，正規分布，混合正規分布について定義する．

(1) 検索語の出現個数を N とし，その出現時間 x_i,$(i = 1,\cdots, N)$ の集合を $X = \{ x_1 , \cdots , x_N \}$ とする．

(2) 正規分布

$$\phi(x;\mu_l,{\sigma_l}^2) = \frac{1}{\sqrt{2\pi{\sigma_l}^2}} \exp\Big(-\frac{(x-\mu_l)^2}{2{\sigma_l}^2}\Big).$$

(3) 混合正規分布

表 1: 混合正規分布による検索語に対する映像区間

混合数	(開始時間, 終了時間)　（単位：分）
1	(9.43,32.15)
2	(16.67,23.28),(22.90,29.77)
3	(16.72,21.92),(18.51,28.85),(23.68,30.58)

$$q_t(x;\theta) = \sum_{l=1}^{m} w_l \phi(x;\mu_l,{\sigma_l}^2),$$

$$\sum_{l=1}^{m} w_l = 1.$$

混合正規分布の混合数を m で表し，混合正規分布のパラメータ $\theta_m = (w_1,\cdots,w_m,\mu_1,\cdots,\mu_m,{\sigma_1}^2,\cdots,{\sigma_m}^2)$，$w_l : l$ 番目の正規分布に対する重み，$\mu_l : l$ 番目正規分布の平均，${\sigma_l}^2 : l$ 番目の正規分布の分散とする．

3.1 EM アルゴリズムによる検索語の分布に対する混合正規分布の近似

映像区間の推定には，VOD 講義の映像に出現する検索語の分布に対して，混合正規分布を近似し，混合正規分布の各正規分布から区間を推定する．混合正規分布の近似は，正規分布の位置 (平均) と幅 (分散) を決める EM アルゴリズムを用い，以下にそのアルゴリズムを示す．

(1) 初期値

μ_l を検索語の出現時間を m 分割した中点，$\sigma_l = 1$, $w_l = \dfrac{1}{m}$ とする．

(2) E-step

$$\eta_{i,l} := \frac{w_l \phi(x;\mu_l,{\sigma_l}^2)}{\sum_{l'=1}^{m} w_l \phi(x;\mu_{l'},{\sigma_{l'}}^2)}.$$

(3) M-step

$$w_l := \frac{1}{n}\sum_{i=1}^{n} \eta_{i,l},$$

$$\mu_l := \frac{\sum_{i=1}^{n} \eta_{i,l} x_i}{\sum_{i'=1}^{n} \eta_{i',l}},$$

$$\sigma_l := \sqrt{\frac{\sum_{i=1}^{n} \eta_{i,l}(x_i-\mu_l)^2}{\sum_{i'=1}^{n} \eta_{i',l}}}.$$

(4)(2),(3) が収束するまで繰り返す．

(5) 区間推定 $\mu_l - \sigma_l$ から $\mu_l + \sigma_l$ までの区間を一つの区間として提供する．

4 混合ベータ分布による映像区間推定

ベータ関数を利用するため，字幕の出現時間 $x \in [0,$終了時刻$]$ を $y \in [0,1]$ に変換して処理を行っている．区間推定の提示の際には，字幕の出現時間に戻している．

(1) ベータ関数

$$B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt.$$

(2) ベータ分布の密度関数

$$f_l(y;\alpha_l,\beta_l) = \frac{y^{\alpha_l-1}(1-y)^{\beta_l-1}}{B(\alpha_l,\beta_l)}.$$

(3) 混合ベータ分布の密度関数

混合正規分布の混合数を t で表し，混合正規分布のパラメータ $\theta_t = (w_1,\cdots,w_l, \cdots,w_t,\alpha_1,\cdots,\alpha_l,\cdots,\alpha_t,\beta_1,\cdots,\beta_l,\cdots,\beta_t)$， w_l：混合数 t の時の l 番目のベータ分布に対する重み，α_l,β_l：混合数 t の時の l 番目ベータ分布のパラメータとするとき，混合ベータ分布の密度関数 $q_t(x;\theta_t)$ を，

$$q_t(y;\theta_t) = \sum_{l=1}^{t} w_l \cdot f_l(y;\alpha_l,\beta_l)$$

で表す．

4.1 EM アルゴリズムによる検索語の分布に対する混合ベータ分布の近似

映像区間の推定には，VOD 講義の映像に出現する検索語の分布に対する近似を，前章の混合正規分布に換えて混合ベータ分布で近似する．また，映像区間の推定についても混合ベータ分布の成分である各ベータ分布から推定する．ベータ分布は，パラメータ α_l と β_l から算出され，それを推定するのに EM アルゴリズムを用い，以下にそのアルゴリズムを示す．

(1) 初期値

t 成分からなる混合ベータ分布の場合，各ベータ分布のパラメータを $\alpha_l = 1$, $\beta_l = 1$,$(l = 1,\cdots,t)$, とする．また，各ベータ分布の重み $w_l = \dfrac{1}{t}$ を初期値とする．

(2) E-step

$$\eta_{i,l} := \frac{w_l f_l(y_i;\alpha_l,\beta_l)}{\sum_{j=1}^{t} w_j f_j(y_i;\alpha_j,\beta_j)}.$$

(3) M-step

$$w_l := \frac{1}{n}\sum_{i=1}^{n} \eta_{i,l},$$

$$\theta_t := \operatorname*{argmax}_{\theta_t} l(\theta_t,\eta).$$

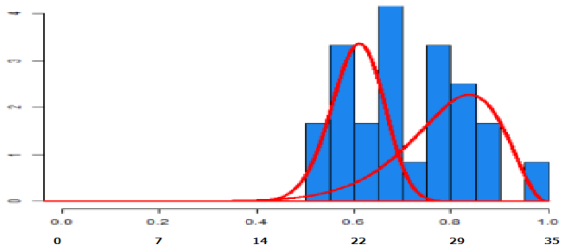

図 2: 混合ベータ分布による近似 (検索語:「広告」,$t=2$)

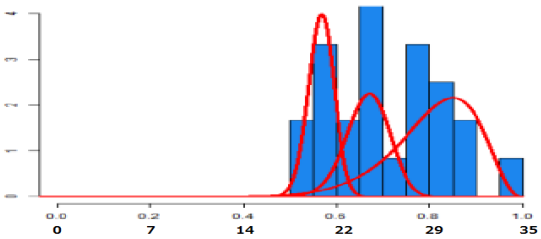

図 3: 混合ベータ分布による近似 (検索語:「広告」,$t=3$)

表 2: 混合ベータ分布による検索語に対する映像区間

混合数	(開始時間, 終了時間)　(単位：分)
1	(20.13,28.90)
2	(18.87,21.34),(23.51,30.07)
3	(17.29,19.04),(20.00,22.94),(24.24,30.17)

ただし，対数尤度関数 $l(\theta_t,\eta)$ は，次式で定義する．

$$l(\theta_t,\eta)=\sum_{i=1}^{n}\sum_{l=1}^{t}\eta_{i,l}\{\log w_l+\log f_l(y_i;\alpha_l,\beta_l)\}$$

(4) (2)(3) を対数尤度関数 $l(\theta_t,\eta)$ が収束するまで繰り返す.

(5) 区間推定

$$\text{分散}\sigma^2_{l,t}=\frac{\alpha_{l,t}\beta_{l,t}}{(\alpha_{l,t}+\beta_{l,t})^2(\alpha_{l,t}+\beta_{l,t}+1)},$$

$$\text{最頻値 } M_{l,t}=\frac{\alpha_{l,t}-1}{\alpha_{l,t}+\beta_{l,t}-2}$$

から映像区間としては，各ベータ分布の密度関数から区間を最頻値 $M_{l,t}$ を中心とした前後 $M_{l,t}-\sigma_{l,t}$ から $M_{l,t}+\sigma_{l,t}$ までを一つの区間として提供する．

5　検索語と共起語の映像区間の推定

検索語に対する共起語の頻度分布の例として，検索語「キーワード」と「サイト」に対する共起語の頻度分布を図 4 に示す.

検索語と共起語の間には関連があると考えられるが，同様に検索語と共起語の頻度分布に対する混合ベータ

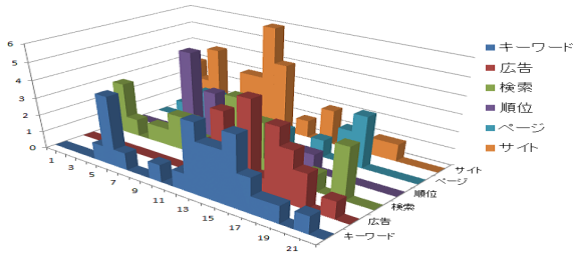

図 4: 共起語の頻度分布 (検索語：「キーワード」)

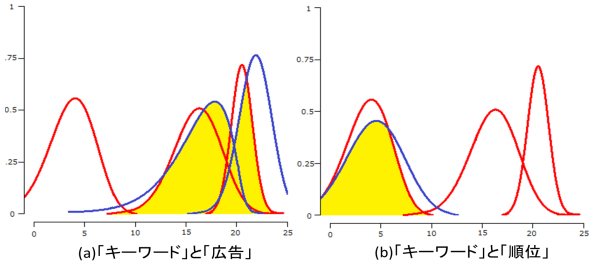

図 5: 検索語と共起語の近似分布の重なり

分布の近似による映像の推定区間の間にも関係があると考えられる．特に，検索語に対する映像の推定区間の中でも話題が相違しているところが測れると考えられる.

そこで，次式で示す検索語の映像の推定区間と共起語の映像の推定区間の一致率と，検索語の映像の推定区間からみた最も一致率の高い区間を求める．ここで，単語 w に対する $t=k$ 個からなる成分の混合ベータ分布の i 番目のベータ分布から得られる区間を $c^i_t(w)$ で表し，同様に w の共起語 c に対する $t=l$ 個からなる成分の混合ベータ分布の j 番目のベータ分布から得られる区間を $I^j_{t=l}(c)$ とする．また，$|I^i_{t=k}(w)|$ を時間，$|I^i_{t=k}(w)\cap I^j_{t=l}(c)|$ を 2 つの区間の共通する時間とする.

$$\text{一致率 } F(I^i_{t=k}(w),I^j_{t=l}(c))=\frac{2}{\frac{1}{\frac{|I^i_{t=k}(w)\cap I^j_{t=l}(c)|}{|I^i_{t=k}(w)|}}+\frac{1}{\frac{|I^i_{t=k}(w)\cap I^j_{t=l}(c)|}{|I^j_{t=l}(c)|}}}$$

検索語に「キーワード」を用いた場合に対する共起語の推定区間と混合ベータ分布の同じ混合数で同じ混合順の推定区間の一致率を表 3 に示す．また，「キーワード」に対する推定区間の最も一致率の高い共起語の推定区間を表 4 に示す.

特に，表 4 では，「キーワード」に対する推定区間が混合ベータ分布の混合数が 3 の時に，混合順が 1 番目の区間に対しては最も共起する共起語では区間が一致せず 3 番目に共起する共起語 3「順位」の映像の推定区間との一致率が高くなっている．一方，2,3 番目の推定区間については，共起語 1「順位」の推定

表 3: 共起語の推定区間の一致率)

		検索語		共起語 1			共起語 2			共起語 3		
混合数	混合順	**キーワード**		**広告**			**検索**			**順位**		
t	l	開始	終了	開始	終了	比率	開始	終了	比率	開始	終了	比率
1	1	16.47	32.09	21.27	30.98	0.81	12.03	30.4	0.83	6.91	20.25	0.26
2	1	13.00	26.29	18.59	22.24	0.64	7.29	20.97	0.59	6.97	10.74	0.0
2	2	21.10	36.13	24.79	32.29	0.75	19.49	36.28	0.95	21.59	24.49	0.60
3	1	6.75	8.10	18.05	19.83	0.0	2.25	10.84	0.58	6.77	7.58	0.80
3	2	18.77	22.94	20.67	24.53	0.57	16.13	20.91	0.48	9.17	11.72	0.0
3	3	23.41	34.09	25.34	32.56	0.84	23.77	35.54	0.92	21.59	24.49	0.24

表 4: 推定区間に対する共起語の最上位推定区間)

		検索語		共起語 1			共起語 2			共起語 3		
混合数	混合順	**キーワード**		**広告**			**検索**			**順位**		
t	l	開始	終了	混合数	混合順	比率	混合数	混合順	比率	混合数	混合順	比率
1	1	16.47	32.09	1	1	0.81	1	1	0.83	2	2	0.59
2	1	13.00	26.29	3	2	0.65	1	1	0.86	2	2	0.61
2	2	21.10	36.13	1	1	0.82	2	2	0.95	2	2	0.60
3	1	6.75	8.10	無	無	-	3	1	0.58	3	1	0.80
3	2	18.77	22.94	2	1	0.89	1	1	0.39	1	1	0.61
3	3	23.41	34.09	2	2	0.85	3	3	0.24	2	2	0.66

区間の一致率が高く，「キーワード」を検索しても違う話題をしている可能性がわかる．検索語「キーワード」と共起語 1「順位」及び共起語 2「広告」の混合ベータ分布の近似分布を図 5 に示す.

6 まとめと今後の課題

本研究では，字幕に対する頻度分布に混合ベータ分布をあてはめ，映像区間の推定を行い，また共起語に対しても同様に映像区間の推定も行った．検索語と共起語の関係が映像区間からも見ることができ，話題を類推する補助ができるのではないかと考えられる．今後は，検索語と複数ある共起語の映像区間から見た関係をより詳しくみるため，検索語と共起語に共起度を導入し，距離を定義することで，全体の話題を推定できるようにしたいと考えている．

参考文献

[1] 北川, 大西, 対面講義と e-learning(LMS + VOD)とを併用した講義形式の実践と分析, 日本教育情報学会学会誌 Vol.22 No.3 pp.57-66, 2007.

[2] Fallon, C. and Brown, S., *e-Learning Standards*, St. Lucie Press, Boca Raton, FL, 2003.

[3] X. Wang, CX. Zhai, and R. Sproat X. Hu., Mining correlated bursty topic patterns from coordinated text streams. In Proc. 13th SIGKDD, pp. 784-793, 2007

[4] V. Lavrenko, M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, J. Allan, Mining of Concurrent Text and Time-Series, KDD-2000 Workshop on Text Mining, 2000.

[5] C. C. Aggarwal, J. Han, J. Wang, and P. S. Yu., On demand classification of data streams. In KDD, pages 503-508, 2004.

[6] A.P.Dempster, N.M.Laird, and D.B.Rubin. *Maximum likelihood form incomlete data via the EM algorithm*, Journal of the Royal Statistical Society series B, Vol. 39, No.1, pp.1-38, 1977.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[8] N. Kobayashi, et al., Detecting Movie Segments Using Gaussian Mixture Models for VOD Lectures with Japanese Subtitles, JSiSE,Vol.10(1), pp.39-46, 2011

[9] Y. Ji, et al., *Applications of beta-mixture models in bioinformatics* Bioinformatics Vol 21 No.9, pp. 2118-2122, 2005.