VOD講義に対する混合正規分布による映像区間の推奨順位の推定

小山 登[1] 羅 毅剛[2] 則本 達哉[3] 椎名 広光[4] 北川 文夫[5]
yamamame@gmail.com[1], evergreen-dc@hotmail.co.jp[2], luoyigang@hotmail.co.jp[3],
shiina@mis.ous.ac.jp[4], kitagawa@mis.ous.ac.jp[5]
岡山理科大学大学院 総合情報研究科[1,2,3]
岡山理科大学 総合情報学部[4,5]

1 まえがき

現在，Web教材を用いたe-Learning，すなわちWBT (Web-Based Training)と対面授業を組み合わせたブレンディッドラーニングやスライドと講義の動画を配信するVOD(Video On Demand)によるe-Learning講義などさまざまな大学でe-Learningの利用が行われている．岡山理科大学ではVODによるe-Learning講義を2004年度から行っている．その中ではVOD教材を見かえすのに，タイトルの内容から目的の教材を探す必要があり，復習が難しい状態にある．

目的の教材VODを検索する方法として，これまで我々は教材として提供されているVODシステムの映像に付加されている字幕を利用して，検索語の頻度のヒストグラムに対して2次の凸関数を当てはめ映像区間を推定する方法[1]を提案し，システムの開発を行った．しかし，この方法では映像区間の数の推定が正しく行うことができない問題があった．また，映像区間の推定に統計的処理方法である混合正規分布モデルの最尤推定をEMアルゴリズムで解く方法[3]を提案した．しかし，この方法では決定した映像区間以外にも映像区間として提供して良いものがあると考えられる．

そこで本研究では，単語頻度分布に対する混合正規分布[4]と同じ単語頻度分布に対するカーネル密度関数[4]との差を用いて，混合正規分布によって推定した映像区間を評価する手法提案する．これにより映像区間を順位付けすることでより良い映像区間を提供する．

2 e-Learning講義システムと検索機能

VODの実行画面は図1(左)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．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．

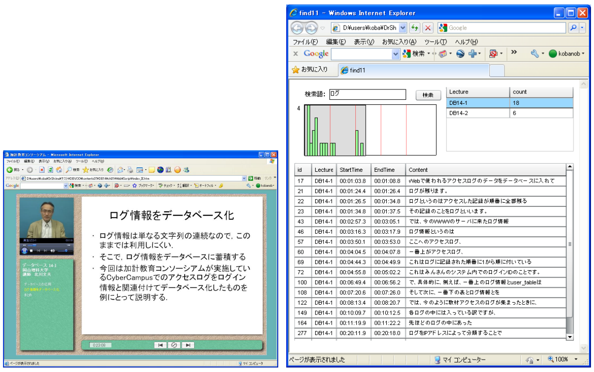

図1: VOD実行画面(左：講義画面，右：検索画面)

これに対して，本研究で追加した機能は，VOD教材の動画の音声データを取り出した字幕データに対して検索語が含まれている頻度をヒストグラムにしたものを利用して検索する．検索画面は図1(右)のような構成で，左上に検索語を入力し，右上のVOD教材(講義回，セクション)を選択すると，その教材に含まれる検索語のヒストグラムと検索語の内容の区間推定の結果を表示する．

3 単語頻度分布対する映像区間推定

本研究においては検索語の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を利用する．このヒストグラムの1つの山を，検索語に関する1つの話題の映像区間であると仮定して，推奨する映像区間を推定する．そこで検索語の出現頻度からヒストグラムの山を推定することになる．しかしながら，山の取り方は1つの山の区間の取り方によって複数種類考えられる．

Copyright(C) 2011 The Association for Natural Language Processing.
All Rights Reserved.

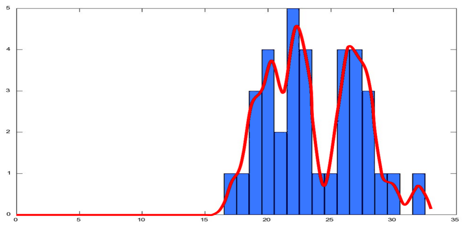

図 2: 単語頻度のヒストグラムとガウス関数による密度分布 (検索語：「広告」)

岡山理科大学サイバーキャンパス 2007 年度データベース 14 回目に対して，14 回目の課題である「インターネットでのビジネスモデルで，キーワード広告が優れている点を論じなさい．また，それ以外に収入が得られそうな方式があるか考えてみよう.」の中から重要語と考えられる「広告」を検索語とした場合を考える．このときのヒストグラムとカーネル密度関数による近似をグラフにしたものを図 2 に示す．図 2 の横軸は映像時間 33 分を 1 分ごとに区切ったものを表し，縦軸にその 1 分間に現れる単語の頻度を表している．このように，山の取り方は 1 つの山の区間の取り方によって複数種類考えることができる．

4 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頻度を近似する．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．

●検索語の出現個数 N とし， その出現時間を $X = \{ x_1 , \cdots , x_N \}$ とする．

●正規分布

$$\phi(x;\mu_l,\sigma_l^2) = \frac{1}{2\pi\sigma_l^2}\exp\Bigl(-\frac{(x-\mu_l)^T(x_i-\mu_l)}{2\sigma_l^2}\Bigr).$$

●混合正規分布

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

混合正規分布の混合数を $m = 1,\cdots,l,\cdots,M$ で表し，混合正規分布のパラメータ $\theta = (w_1,\cdots,w_m,\mu_1,\cdots,\mu_m,\sigma_1^2,\cdots,\sigma_m^2)$，$w_l$：$l$ 番目の正規分布に対する重み，μ_l：l 番目正規分布の平均，σ_l^2：l 番目の正規分布の分散とする．

4.1 映像区間推定処理

映像区間推定処理は正規分布の位置を決める EM アルゴリズムと，開始と終了の区間推定の順で行われる．

[EM アルゴリズム]
混合正規分布の各パラメータは EM アルゴリズムによってよって決定する．

●初期値

μ_l を検索語の出現時間のを m 分割した中点，$\sigma_l = 1$，$w = 1$ とする．

● E-step

$$\eta_{i,l} := \frac{\hat{w}_l\phi(x;\hat{\mu}_l,\hat{\sigma}_l^2)}{\sum_{l'=1}^{m} w_l\phi(x;\hat{\mu}_{l'},\hat{\sigma}_{l'}^2)}.$$

● M-step

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

$$\hat{\mu}_l := \frac{\sum_{i=1}^{n}\hat{\eta}_{i,l}x_i}{\sum_{i'=1}^{n}\hat{\eta}_{i',l}},$$

$$\hat{\sigma}_l^2 := \sqrt{\frac{\sum_{i=1}^{n}\hat{\eta}_{i,l}(x_i-\hat{\mu}_l)^2}{\sum_{i'=1}^{n}\hat{\eta}_{i',l}}}.$$

[区間推定処理]
単語の出現時間に対して混合正規分布で EM アルゴリズムによって出現頻度を近似する．このとき，一つのセクションは 20〜30 分であるので検索語に対する映像区間は 5 程度であるとし，混合正規分布の混合数 $m = 1,\cdots,5$ で近似を行う．これによって得られた各正規分布に対して，68.26%をカバーする平均 μ_l から標準偏差 $\pm\sigma_l$ の幅を区間とする．推定されたパラメータ $\hat{\theta} = (\hat{w_1},\cdots,\hat{w_m},\hat{\mu_1},\cdots,\hat{\mu_m}\ ,\hat{\sigma_1^2},\cdots,\hat{\sigma_m^2})$ に対して，$\hat{\mu_l}-\hat{\sigma_l}$ から $\hat{\mu_l}+\hat{\sigma_l}$ までの区間を一つの区間として提供する．

5 評価値による推定映像区間の順位付け

映像区間の順位付けにはガウス関数を用いたカーネル密度関数と混合正規分布の差を用いて，評価値をつけることによって行う．また，映像区間を評価するために次の式による m 個の正規分布のうち k 番目を取り除いた正規分布と，取り除かない分布の比較によって各正規分布を評価する．

$$q'_{t,k}(x,\theta) = \sum_{l=1,l\neq k}^{m} w_l\phi(x;\mu_l,\sigma_l^2).$$

Copyright(C) 2011 The Association for Natural Language Processing.
All Rights Reserved.

5.1 カーネル密度関数による単語頻度の近似

カーネル密度関数のカーネル関数としては一般的なガウス関数を用いる．カーネル密度関数は以下に定義する．

●検索語の出現個数 N とし， その出現時間を $X = \{ x_1 , \cdots , x_N \}$ とする．

● h：全映像時間の幅とする．

● b：一つのデータが持つバンド幅とする．本研究では検索語の影響は出現時間からから前後 1.5 分としている．そのため，データから前後 1.5 分の幅の分布を作る $b = 0.75$ とする．

●カーネル関数

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2).$$

●カーネル密度関数

$$KD_t(x) = \frac{1}{N \cdot b} \sum_{i=1}^{N} K\Big(\frac{x - x_i}{h}\Big).$$

5.2 順位付けのための推定映像区間の評価

各推定映像区間の評価の計算は，m 個の線形結合による混合正規分布に対して，評価する映像区間に対応する k 番目の正規分布を取り除いて他の正規分布 $m-1$ 個を線形結合した分布との差から決定する．この評価は取り除いた分布が全体の分布に対する影響度合いを評価すること目的としている．本研究では映像区間の評価に差の二乗和，カルバックライブラー情報量による差，混合正規分布のカルバックライブラー情報量の 3 つの式を用いて評価する．なお，混合数 $m = 1$ のときには，評価区間の差を評価できないので例外とする．

(評価 1) 差の二乗和

カーネル密度関数を真の確率分布として，カーネル密度関数と推定した混合正規分布との差を映像区間である $\mu - \sigma$ から $\mu + \sigma$ の秒ごとの二乗和を評価する．

$$V_1 = \sum_{t=\mu-\sigma}^{\mu+\sigma} \Big(\big(KD_t(x) - q_t(x,\theta)\big)^2 \Big) - \sum_{t=\mu-\sigma}^{\mu+\sigma} \Big(\big(KD_t(x) - q'_{t,k}(x,\theta)\big)^2 \Big).$$

(評価 2) カルバック・ライブラー情報量による差

カーネル密度関数を真の密度分布として，カーネル密度関数と推定した混合正規分布との差を映像区間である $\mu-\sigma$ から $\mu+\sigma$ の秒ごとのカルバック・ライブラー情報量を用いて評価する．

$$V_2 = \sum_{t=\mu-\sigma}^{\mu+\sigma} \Big(KD_t(x) \log \frac{KD_t(x)}{q_t(x,\theta)} \Big) - \sum_{t=\mu-\sigma}^{\mu+\sigma} \Big(KD_t(x) \log \frac{KD_t(x)}{q'_{t,k}(x,\theta)} \Big).$$

(評価 3) 混合正規分布のカルバック・ライブラー情報量

推定した混合正規分布と評価する分布を取り除いた混合正規分布との差を，全区間に対して秒ごとに，直接，カルバック・ライブラー情報量を用いて評価する．

$$V_3 = \frac{1}{2} \Bigg(\sum_{t=0}^{n} \Big(q_t(x,\theta) \log \frac{q_t(x,\theta)}{q'_t(x,\theta)} \Big) + \sum_{t=0}^{n} \Big(q'_t(x,\theta) \log \frac{q'_{t,k}(x_t,\theta)}{q_t(x,\theta)} \Big) \Bigg).$$

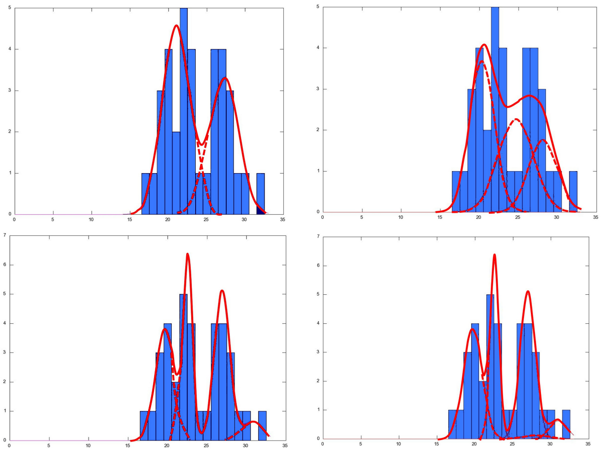

図 3: 混合正規分布による近似 (検索語：「広告」, 左上:$m = 2$, 右上:$m = 3$, 左下:$m = 4$, 右下:$m = 5$)

表 1: 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	(9.43,35.90)
2	(16.67,23.28),(22.90,29.77)
3	(16.72,21.92),(18.51,28.85),(23.68,30.58)
4	(17.16,20.28),(21.25,21.95),(24.91,26.94),(28.20,31.37)
5	(17.16,20.28),(21.25,21.95),(23.68,30.23),(24.91,26.96),(28.67,31.30)

6 評価実験

検索語に「広告」を対する区間推定の処理例を示す．

(1) 混合正規分布による近似

検索語「広告」の開始からの出現時間 $X = \{17.25, 13.31, \cdots, 32.00\}$ を，混合正規分布を混合数 $m = 1, \cdots, 5$ によって単語の出現頻度を近似する．近似によって得られた分布を図 3 に示す．

(2) 区間の推定

Copyright(C) 2011 The Association for Natural Language Processing.
All Rights Reserved.

表 2: 評価 1(V_1) による順位

混合数	映像区間の出現順				
2	1	2			
3	3	8	11		
4	6	9	4	12	
5	7	10	14	6	13

表 3: 評価 2(V_2) による順位

混合数	映像区間の出現順				
2	1	4			
3	7	8	9		
4	2	10	5	12	
5	3	11	14	6	13

各正規分布の平均と分散から映像区間を求める．検索語「広告」の推定した映像区間を表 1 に示す.

(3) 推定された区間の順位

推定された分布を評価値によって順位付けする．それぞれの評価法での混合数 $m = 1$ を除く映像区間の中で順位を表 2，表 3，表 4 に示す.

7　評価

検索語「広告」に対する推定した映像区間の順位表 2, 3，4 において混合数 $m = 4$ のときの 1 つ目の分布から作られる映像区間は 評価 1 の評価値では順位が他の評価値に比べて低く，評価 2 と 評価 3 では $m = 2$ の 1 つ目の分布と順位が入れ替わっている．これは検索語が出現する時間の前半を 1 つの山と処理するか，2 つの山と処理するかで評価が変化する個所である.

評価値による順位を人手で作成した正解順位と比べたときの差の最大値，平均，分散を表 5 に示す．結果から人手での順位との差が少ない評価値は 評価 1 である. しかし，分散は最も大きく，値のばらつきが多い評価となっていることが分かる．最大値と分散が最も小さい評価値は 評価 2 である．本研究では最大値と分散が最も小さいことから，評価 2 の評価値が他の検索語にも有効であると考える.

また，順位付けの結果を人手で判断すると，すべての順位から検索語「広告」の場合 評価 1 では 5 位まで, 評価 2，評価 3 では 6 位まで提供できる区間である.

8　まとめ

字幕を用いた VOD 教材検索システムの提案を行った. 提案した単語頻度分布による検索システムは，ヒストグラムの山を EM アルゴリズムを用いて混合正規分布で近似する方法を用いている．得られた分布から映像区間を推定し順位づけを行った．先行研究 [2] での，AIC を元に得られた映像区間以外にも提供できる映像区間が存在すると考えたためである．順位付けに対して，3 つの評価値を比較検討した.

表 4: 評価 3(V_3) による順位

混合数	映像区間の出現順				
2	2	4			
3	7	10	12		
4	1	8	5	11	
5	3	9	14	6	13

表 5: 評価法による順位と人手による順位の差

評価法	差の最大	平均	分散
評価 1(V_1)	6	2.53	9.55
評価 2(V_2)	5	3.47	7.41
評価 3(V_3)	7	3.34	9.38

単語の出現頻度に対する映像区間の検索は，人間の感覚の映像区間とは違いがあるかもしれない．しかし，大学の講義などのように，ある一定の目的をもった動画に対しては有効な検索方法であると考えている.

参考文献

[1] 椎名，小林，北川，VOD 講義の字幕を利用した話題検索，pp547-550, 言語処理学会第 16 回年次大会発表論文集, 2010.

[2] 小林, 椎名, 北川, 字幕データ付き VOD 講義の単語頻度に対する混合正規分布モデルによる映像区間の推定, pp306-307, 日本教育情報学会大 26 回年会,2010.

[3] 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

[4] 金森，竹ノ内，村田，パターン認識，共立出版, 2009.

[5] 伊藤, 藤井, 石川, 音声文書検索を用いたオンデマンド講義システム, 電子情報通信学会技術研究報告 SP 音声, Vol.101, No.523, pp.55-60, 2001.

Copyright(C) 2011 The Association for Natural Language Processing. All Rights Reserved.