

VOD 講義の字幕を利用した話題検索

椎名広光¹小林伸行²北川文夫¹

shiina@mis.ous.ac.jp koba_nob@sguc.ac.jp kitagawa@mis.ous.ac.jp

岡山理科大学 総合情報学部¹ 山陽学園大学 総合人間学部²

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 [4] 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. 岡山理科大学では関連 6 大学で教育コンソーシアムを構成し, 単位互換制度を利用した VOD による e-Learning 講義の相互提供を 2004 年度から行っている [3]. その中では学習者が多くの VOD 教材を復習する場合, タイトルやスライドの内容から目的の教材を探す必要がある, これまで VOD 教材の検索方法としては講義のスライド内 (Microsoft PowerPoint) や教師の音声からもスライドの検索が可能なシステムが開発されている [2]. しかし, そのシステムでは, 語句が含まれているスライドが羅列されるだけであり, 講義のどのあたりでどのくらいの時間が検索語に関する内容が扱われているのかわからないという問題がある. また, スライドの作り方に依存性が高く, アニメーションを多用する事例には不向きと考えられるほかに, 音声認識の誤認識の問題が含まれるのではないかと考えられる.

本研究では, 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, 検索単語の頻度のヒストグラムを提示や検索語に関連する映像の区間を推定を行うことで, 効率的な学習を目的としている. また, これはスライドの作成方法に依存しないような VOD 教材にでも対応するためである.

作成しているシステムの処理としては, 検索にはキーワード検索のように単語を用い, 映像を時間単位の単語頻度をヒストグラムにし, 次にこのヒストグラムの分布の 1 つの山を検索語の関連した映像区間であると仮定して, 推奨する映像区間を推定している. 映像検索については, 単語の出現と代数的アプローチ [1] が知られているが, 単語の頻度分布に対して, 統計的処理を組み合わせるアプローチを試みている.

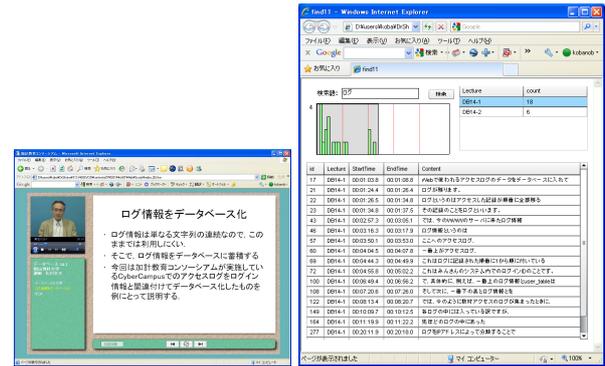


図 1: VOD 実行画面 (左: 講義画面, 右: 検索画面)

2 VOD システムによる 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.

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 検索語の頻度による比較

3.1 字幕に対する検索語の頻度

本研究では現在教材として提供されている VOD システム [3] の映像に付加されている字幕に検索を行い, 映

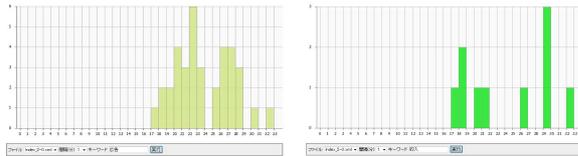


図 2: 単語頻度 (左:「広告」, 右:「収入」)

像区間を推定する．そのため，検索語の頻度を時間間隔ごとに表示したヒストグラムが必要となる．

本稿の例では，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 回目の VOD 教材に対して，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」と「収入」を検索語とした場合，単語頻度のヒストグラム (図 2) が作成できる．図 2 の場合，横軸は映像時間 33 分を 1 分ごとに区切った横軸とその 1 分間に現れる単語の頻度を縦軸に表したものである．

4 字幕に対する単語頻度分布による映像区間推定

4.1 字幕に対する単語頻度分布による映像区間推定の問題

字幕に対する単語頻度から作られるヒストグラムの 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．そこでヒストグラムから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ることができる．例えば，検索語を「広告」としたヒストグラム (図 2) に対する場合では，山を最も細かく考えて山が 4 つと推定する図 3(左)，山を最も大きく考えて山が 1 つと推定すると図 3(右)，山をある程度幅を持たせと考えると，山を 2 つと推定する図 4(左) と図 4(右) の 4 種類の候補を考えることが順等で，この中から 1 つを利用者に推奨したり選択してもらう．

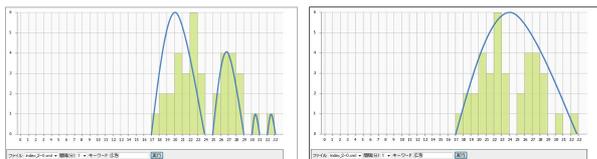


図 3: 単語分布に対する二次関数の近似 1 (左:4 区間, 右:1 区間)

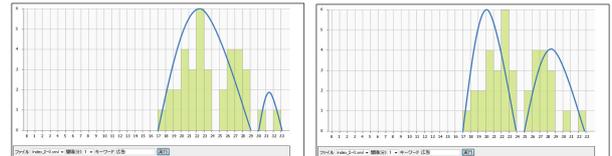


図 4: 単語分布に対する二次関数の近似 2 (左:2 区間, 右:2 区間)

4.2 字幕に対する単語頻度分布による映像区間推定の手法

字幕に対する単語頻度から作られるヒストグラムの 1 つの山の推定には，単語頻度が連続して出現しているヒストグラムと上に凸の二次関数を比較してその当てはまり具合を用いる．また，複数の山の推定は，初めの山の区間をクラスタリング [3] の手法と同様に近い山を統合する．

[定義]

映像時間 $T = 0..n$ ，単語の頻度を計算する 1 つの間隔を t_i とすると，映像時間 $T = t_0 t_1 \dots t_n$ と表し，本研究で利用している 1 分間隔の場合では， $t_0 = 0..1$ ， $t_1 = 1..2$ ， \dots ， $t_{n-1} = n-1..n$ となる．

$C_W(t_i)$: 検索単語 W に対する区間時間ごとの単語出現頻度．映像区間 T に対しては， $C_W = \{C_W(t_0), C_W(t_1), \dots, C_W(t_{n-1})\}$ で表す．

区間統合処理の繰り返し i 番目における区間 $M^i = \{M_0^i, M_1^i, \dots, M_{n_i-1}^i\}$ ． M^i の要素 $M_k^i = a..b$ は，ヒストグラムの一つの山に対応する区間時間とする．

区間 M^i を求めるための候補となる区間 $M^{i,j} = \{M_0^{i,j}, M_1^{i,j}, \dots, M_{n_i-1}^{i,j}\}$ ． $M^{i,j}$ の要素 $M_k^{i,j} = a..b$ は，ヒストグラムの一つの山に対応する区間時間とする．

評価値 V^i : 区間 M^i に対する評価値．また，区間 M^i の可能性のある区間 $M^{i,j}$ に対する評価値を評価値 $V^{i,j}$ とする．

[区間推定処理]

(Step1) 単語頻度が連続して表れている区間を求め，その区間ごとに上に凸の二次関数をあてはめ，そのあてはまり具合を評価する．

(Step1.1) 最初に検索単語 W に対する時間区間ごとの単語出現頻度 $C_W = \{C_W(t_0), C_W(t_1), \dots, C_W(t_{n-1})\}$ から単語が連続して出現する区間のみを出現順に集めて，区間 $M^1 = \{M_0^1, M_1^1, \dots, M_{n_1-1}^1\}$ を求める．この区間は， n_1 個の区間数を表し，検索語が表れていない区間を除くため $0..n$ まで時間のうち一部が取り出されることになる．

(Step1.2) 最初の区間 $M^1 = \{M_0^1, M_1^1, \dots, M_{n_1-1}^1\}$

の各区間 M_k^1 に対して, 上に凸の二次関数を当てはめる.

(1) 区間 $M_k^1 = a..b$ の最大頻度を

$$c = \max_{l=a}^{b-1} C_W(t_l^0)$$

として, 次の上に凸の二次関数 $B(x)$ を次式で当てはめることにする.

$$B(x) = \frac{-4c}{(b-a+1)^2}(x-a+1)(x-b)$$

(2) 区間 M_k^1 と上に凸の二次関数との差を当てはまり具合を $V_k^1(t_k^1)$ とする.

$$V_k^1(t_k^1) = \sum_{l=a}^{b-1} |C_W(t_l) - B(l)|$$

(Step1.3) 区間全体をを評価する. そのために, 複数の山の区間に対応する区間 M^1 に対して評価 $V(t^1)$ を次の式で求める.

$$V^1 = \frac{1}{\sum_{k=0}^{n_1-1} V(t_k^{i,j})}$$

(Step2) 区間 M^{i-1} から隣接する区間を統合後, 最も良い区間 M^i を求めるために, $n_i - 1$ 種類の区間の候補を求める. なお, Step2 の一回目のみの繰り返し用変数 i を $i = 2$ としておく.

このステップの処理は, $j = 0..n_{i-1} - 1$ の1つに対して, M^{i-1} 内の隣接する区間 M_j^{i-1} と M_{j+1}^{i-1} を $M_j^{i,j}$ に統合し, 区間 $M^{i,j} = \{M_0^{i,j}, M_1^{i,j}, \dots, M_{n_i-1}^{i,j}\}$ を作成し, 次の Step2.1 と Step2.2 を繰り返す.

(Step2.1) 区間 $M^{i,j} = \{M_0^{i,j}, M_1^{i,j}, \dots, M_{n_i-1}^{i,j}\}$ の各区間 $M_k^{i,j}$ を上に凸の二次関数を当てはめる.

(1) 区間 $M_k^{i,j} = a..b$ の最大頻度を

$$c = \max_{l=a}^{b-1} C_W(t_l)$$

とすると, 次の上に凸の二次関数 $B(x)$ を次式で当てはめることにする.

$$B(x) = \frac{-4c}{(b-a+1)^2}(x-a+1)(x-b)$$

(2) 区間 $M_k^{i,j}$ と上に凸の二次関数との差をあてはまり具合 $V_k^{i,j}(t_k^{i,j})$ とする.

$$V_k^{i,j}(t_k^{i,j}) = \sum_{l=a}^{b-1} |C_W(t_l) - B(l)|$$

表 1: 検索語の区間評価

時間	広告	近似値	各区間の評価値
17..18	1	2.625	1.625
18..19	2	4.5	2.5
19..20	2	5.625	3.625
20..21	4	6	2
21..22	3	5.625	2.625
22..23	6	4.5	1.5
23..24	3	2.625	0.375
24..25	0	0	-
25..26	3	2.56	0.44
26..27	4	3.84	0.16
27..28	4	3.84	0.16
28..29	3	2.56	0.44
29..30	0	0	-
30..31	1	1	0
31..32	0	0	-
33..34	1	1	0

(Step2.2) 複数の上に凸の山を評価する. そのために, 複数の上に凸の山の区間に対応する区間 M^i に対して評価 $V^{i,j}(t^{i,j})$ を次の式で求める.

$$V^{i,j}(t^{i,j}) = \frac{1}{\sum_{k=0}^{n_i-1} V(t_k^{i,j})}$$

(Step3) 繰り返し i 番目における複数の区間の候補 $M^{i,0}M^{i,1}..M^{i,n_i-1}$ に対する評価 $V^{i,0}(M^{i,0}), V^{i,1}(M^{i,1}), \dots, V^{i,n_i-1}(M^{i,n_i-1})$ の内, 最大となる区間の候補を最も良い区間 M^i とする.

(1) 評価最大の区間の順序の取得 $k = \operatorname{argmax}_{j=0}^{n_i-1} V(t^{i,j})$

(2) 評価最大の区間 $M^i = M^{i,k}$

(Step4) $i = i + 1$ とし, Step2 と Step3 での区間数が利用者の判断によって停止されるまで繰り返す.

4.3 計算例

検索語に「広告」を利用した場合の区間推定の処理例を示す.

Step1 では, 単語の頻度分布から単語が連続して現れる区間が作られる. 「広告」の場合は, 図 3(右) のように $M^1 = \{17..24, 25..29, 30..31, 32..33\}$ の4つの区間がある. これに対して, 単語頻度と二次関数の近似値とから区間の評価値(表1)は, それぞれ以下のとおりである.

$$V_0^1(17..24) = 1.625 + 2.5 + 3.625 + 2 + 2.625 + 1.5 + 0.375 = 14.25$$

$$V_1^1(25..29) = 0.44 + 0.16 + 0.16 + 0.44 = 1.2$$

$$V_2^1(30..31) = 0$$

$$V_3^1(32..33) = 0$$

この4区間全体の評価値は

$$V^1(17..33) = \frac{1}{15.45} = 0.0647$$

となる。

4.2節のStep2 ($i = 2$)では、区間の候補は次の3種類ある。

$$(1)M^{2,0}=\{17..29,30..31,32..33\}$$

$$(2)M^{2,1}=\{17..24,25..31,32..33\}$$

$$(3)M^{2,2}=\{17..24,25..29,30..33\}$$

これに対して、区間の候補内の1区間それぞれの評価は

$$(1)V_0^{2,0}(17..29)=21.107, V_1^{2,0}(30..31)=0, \\ V_2^{2,0}(32..33)=0$$

$$(2)V_0^{2,1}(17..24)=14.25, V_1^{2,1}(25..31)=7, \\ V_2^{2,1}(32..33)=0$$

$$(3)V_0^{2,2}(17..24)=14.25, V_1^{2,2}(25..29)=1.2, \\ V_2^{2,2}(30..33)=1.5$$

と計算することができ、区間の候補の評価値は、次のとおりである。

$$(1)V^{2,0}(17..33)=\frac{1}{21.107}=0.0474$$

$$(2)V^{2,1}(17..33)=\frac{1}{21.5}=0.0471$$

$$(3)V^{2,2}(17..33)=\frac{1}{16.95}=0.0590$$

このうち $V^{2,2} = 0.0590$ が最大となり、 $M^{2,1}$ を最も良い区間とする。

4.2節のStep2 ($i = 3$)では、区間の候補は次の2種類ある。

$$(1)M^{3,0} = \{17..31, 32..33\}$$

$$(2)M^{3,1} = \{17..24, 25..33\}$$

これに対して、区間の候補内の1区間それぞれの評価は

$$(1)V_0^{3,0}(17..31) = 21.107, V_1^{3,0}(32..33) = 1.5$$

$$(2)V_0^{3,1}(17..24) = 14.25, V_1^{3,1}(25..33) = 13.901$$

と計算することができ、区間の候補の評価値は、次のとおりである。

$$(1)V^{3,0}(17..33) = \frac{1}{22.607} = 0.0442$$

$$(2)V^{3,1}(17..33) = \frac{1}{28.151} = 0.0355$$

このうち $V^{3,0} = 0.0442$ が大きいので $M^{3,0}(17..33)$ が最も良い区間とする。

4.2節のStep2を $i = 4$ まで実行した場合、区間の候補は、

$$M^{4,0} = \{17..33\}$$

と計算することができ、区間の候補の評価値は、次のとおりである。

$$V^{4,0}(17..33) = \frac{1}{31.803} = 0.0314$$

となる。

最後に4.2節のStep2 ($i = 3$)での2種類の区間の候補 $M^{3,0} = \{17..31, 32..33\}$, $M^{3,1} = \{17..24, 25..33\}$ は、それぞれ図4の左と右に対応する。また、評価値が大きい区間 $M^{3,1} = \{17..24, 25..33\}$ は、図4左の区間の方がよさそうであることと一致する。

5 まとめ

字幕データを用いた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、ヒストグラムの山の上に凸の二次関数を用いて近似する方法を用いている。これを用いたのは計算が容易で速いためである。また、離れすぎた2つの区間を統合する場合には、評価値が降下するように作られている。そのため話題が途切れている区間を統合することは少ないようになっている。ヒストグラムの山に対する近似については、話題のなかでの単語の出現が人間の感覚と相違しているかもしれない。しかしながら、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、検索するの有効であろうと考えている。

参考文献

- [1] 西尾, 田中, 上原, 有木, 加藤, 河野, 情報の構造と検索, 岩波講座, マルチメディア情報学, 2000.
- [2] 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.
- [3] 北川, 大西荘一 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.
- [4] 田中他, 学用語教育における e-learning: ブレンディッド・ラーニングの実践と評価, 川崎医療福祉学会誌, Vol.17, No.1, pp.153-162, 2007.
- [5] 小林, 椎名, 北川, 字幕データを用いたVOD教材検索システムの提案, pp416-417, 教育情報システム学会第31回全国大会, 2009.