

バースト検出に基づく映像からのトピック抽出

白浜 公章[†], 上原 邦昭^{††}

[†] 神戸大学大学院経済学研究科 ^{††} 神戸大学大学院工学研究科

映像という連続メディアには、「意味内容が時区間で継続する」という特徴がある。そこで、本論文では、映像を意味内容によってラベル付けされた時区間からなる“インターバルシーケンス”としてモデル化する。とりわけ、プロによって編集されたフィクション映像を対象として、登場人物の出現区間と非出現区間を表わすインターバルシーケンスを作成する。そして、“バースト”と呼ぶ異常な編集技法を検出する手法を提案する。具体的には、確率モデルに基づく時系列セグメンテーション手法を用いて、出現区間長が異常に短くなる、もしくは長くなるパターンをバーストとして検出する。最終的に、バーストによって特徴づけられた映像区間では、視聴者にインパクトを与える“トピック”が表現されていることを示す。

Topic Extraction in Videos Based on Burst Detection

Kimiaki SHIRAHAMA[†] Kuniaki UEHARA^{††}

[†] Graduate School of Economics, Kobe University

^{††} Graduate School of Engineering, Kobe University

A video is a continuous media where the same semantic contents continue in time intervals. So, we model it as an “interval sequence” consisting of intervals labeled by semantic contents. In particular, we target fiction videos edited by professional editors and model them as interval sequences consisting of intervals of character's appearance and disappearance. From such sequences, we detect “bursts” as abnormal editing techniques. Specifically, by using probabilistic time series segmentation method, we detect bursts as abnormally shortening or lengthening patterns of appearance intervals. As a result, subsequences characterized by bursts correspond to “topics” which have much impacts on viewers.

1 はじめに

本論文では、映像データに対する異常検出手法の1つとして、プロの編集者によって編集された映像から“異常な編集技法”を検出する手法を提案する。ここで、“異常な編集技法”が使用される映像区間は、視聴者にインパクトを与える“トピック”に対応していると考えられる。例えば、サスペンス映画の場合、ショットの切り替わりが異常に速い映像区間では、「殺人」や「逃走」といった、映画の中でも特にスリリングな意味内容が表現されている。したがって、異常な編集技法の検出は、印象的なトピックの発見につながることになる。

2 映像のモデル化とバースト

映像は、フレームが時間軸上で連続的に再生されて、意味内容を伝達する“連続メディア”である¹⁾。すなわち、映像には、「意味内容が時区間で継続する」という特徴がある。そこで、映像を、意

味内容によってラベル付けされた時区間からなる“インターバルシーケンス”としてモデル化する。とりわけ、ショット内では、1つのまとまりのある意味内容が継続していると仮定して、インターバルシーケンスを作成する。

一般に、フィクション映像では、主人公、ヒロイン、悪役といった登場人物を軸としてストーリーが進展して行く。プロの編集者は、「新しい意味を付け加えるアクションだけをスクリーンに映し、それ以外の冗長なアクションは排除する」という原則に従って、ショットを選択している²⁾。その結果、ショット中の登場人物の“出現”や“非出現”には、映像の意味内容が強く反映される。ここで、ショットの切り替わりに伴って、画面上に出現する登場人物は異なってくる。例えば、図1の *shot 1* では女性の登場人物Aと2人の男性B, Cの3人、*shot 2* ではBのみ、*shot 3* ではAのみが出現している。今、Aの出現と非出現に基づいてショットをラベ

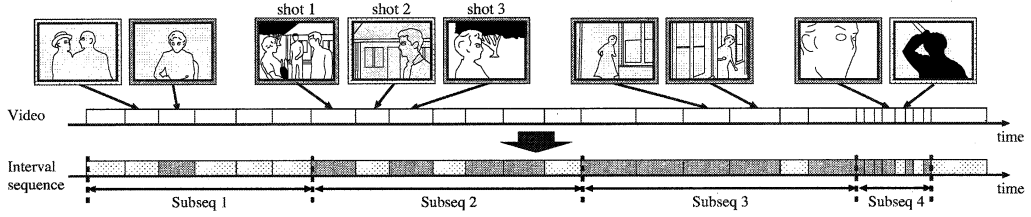


Fig. 1 インターバルシークエンスとしての映像のモデル化

ル付けすれば、図1下部にあるインターバルシークエンスが作成できるものとする。ここで、 A は、濃く網掛けされた時区間に出現し、薄く網掛けされた時区間には出現していない。

上記のインターバルシークエンスにおいて、*Subseq 3*では、 A の出現区間が大半を占めていることが分かる。これは、 A があちこち動き回っており、 A の単独のアクションを様々なカメラから撮影しているからである。一方、*Subseq 2*では、 A の出現区間と非出現区間がほぼ交互に繰り返されている。これは、 A と他の人物との会話が、各自1人を映したショットを交互につなぎ合わせて表現されているからである。また、*Subseq 1*では、 A の非出現区間が大半を占めている。これは、 A が他の人物の会話を聞いているだけで、ストーリーの進行上重要でなくなっているからである。さらに、*Subseq 4*では、 A の出現区間長が非常に短くなっている。これは、 A が殺害される様子が、時間長の短いショットを連続的につなぎ合わせて表現されているからである。このように、登場人物の出現区間と非出現区間からなるインターバルシークエンスには、意味内容の特徴づける様々な出現パターンが含まれている。

インターバルシークエンスには、意味内容に応じた時区間の局所性が存在する。これは、“モードに相関した編集のリズム”という編集技法によって裏付けられる⁴⁾。一般に、スリリングなモードは高速なショットの切り替え(図1の*Subseq 4*)、ロマンティックなモードはゆっくりとしたショットの切り替えによって演出されることが知られている。このようなショットの時間長の局所性を、「出現区間長の異常な短縮もしくは延長」という異常パターン(“バースト”)として検出する。最終的に、上記のような編集技法を考慮すると、バーストが発

生している映像区間では、視聴者にインパクトを与えるトピックが表現されていると考えられる。

3 バースト検出手法

本手法では、まず、インターバルシークエンスを、出現パターンが類似している“サブシークエンス”に分割する。とりわけ、図1に示されているように、出現区間長、非出現区間長、出現区間と非出現区間の発生率という3つの観点から、出現パターンの類似性を評価する。そして、サブシークエンスごとにバーストが発生しているか検証してトピックを抽出する。例えば、図1の*Subseq 4*では、出現区間長が異常に短くなるバーストが発生しているため、トピックとして抽出される。

まず初めに、インターバルシークエンス X は、以下のように定式化できる。

$$X = x_1, x_2, x_3, \dots, x_N \quad x_i = (l_i, v_i). \quad (1)$$

ここで、 i 番目の時区間 x_i に対して、 $l_i \in \mathbb{R}$ は区間長、 $v_i \in \{A(\text{pppearance}), D(\text{isappearance})\}$ は出現もしくは非出現というラベルを表わしている。

次に、時系列セグメンテーション手法⁵⁾を用いて、 X を互いに重ならない $K (\ll N)$ 個のサブシークエンス S に分割する。すなわち、 S は以下のように定式化できる。

$$S = s_1, s_2, s_3, \dots, s_K \quad s_i = x_a, x_{a+1}, \dots, x_b. \quad (2)$$

ここで、 s_i は a 番目から b 番目までの時区間から構成されている。さらに、以下の確率モデルを用いて、 s_i の出現パターンの類似性を評価する。

$$\begin{aligned} p(s_i) &= \prod_{j=a}^b p(x_j) \\ &= \prod_{j=a}^b p_A(l_j) \cdot p(v_j = A) \quad \text{if } v_j = A \\ &\quad p_D(l_j) \cdot p(v_j = D) \quad \text{if } v_j = D. \end{aligned} \quad (3)$$

上式では、 s_i 内の時区間 $x_j = (l_j, v_j)$ が観測される確率 $p(s_j)$ を 3 つの確率分布から算出している。まず、 $p(v_j)$ は、ラベル v_j の確率分布であり、出現 ($v_j = A$) もしくは非出現 ($v_j = D$) が観測される確率を表す。すなわち、 s_i に含まれる時区間のラベルが、単一の確率分布 $p(v_j)$ に従っているか検証して、出現区間と非出現区間の発生率の一致性を評価している。また、 $p_A(l_j)$ と $p_D(l_j)$ は、それぞれ出現区間長と非出現区間長の確率分布である。つまり、これらの確率分布を用いて、 s_i における出現区間長、及び非出現区間長の類似性を評価している。最終的に、上記の確率分布に基づいて、 s_i 内の全ての時区間が観測される同時確率を $p(s_i)$ とする。ゆえに、 $p(s_i)$ は、 X における a 番目から b 番目までの時区間を s_i にまとめた際の総合的な評価値を表わしている。

上記の確率モデルを踏まえて、インターバルシークエンス X を K 個のサブシークエンスに分割する。これは、以下の式によって、 X 内の全ての時区間が最も高確率で観測できる K 個のサブシークエンスを求める問題に他ならない。

$$P(X) = \prod_{i=1}^K p(s_i). \quad (4)$$

ただし、計算を容易にするために、 $P(X)$ の自然対数をとって、 $P'(X) = \log(P(X)) = \sum_{i=1}^K \log p(s_i)$ を最大化させている。ここで、自然対数は単調増加関数であり、 $P(X)$ を最大化することは、 $P'(X)$ を最大化することと同値である。最終的に、動的計画法を用いれば、 $P'(X)$ を最大化する最適な K のサブシークエンスが求められる。

以下の評価尺度を用いて、サブシークエンス s_i における出現区間長の異常性（“バースト性 (Burstiness (B))”）を評価する。

$$B(s_i) = \frac{T_A^{s_i}}{T^{s_i}} \times \int_0^\infty |\lambda_A^{s_i} e^{-\lambda_A^{s_i} x} - \bar{\lambda}_A e^{-\bar{\lambda}_A x}| dx. \quad (5)$$

ここで、 T^{s_i} は s_i の時間長、 $T_A^{s_i}$ は s_i 内の出現区間長の合計を表わしている。すなわち、第 1 項を用いて、登場人物がほとんど出現していない映像区間は、不適切にトピックとして抽出されないようにしている。また、第 2 項では、 s_i 内の出現区間から推定された指数分布 ($\lambda_A^{s_i} e^{-\lambda_A^{s_i} x}$) と、映像全体での出現区間から推定された指数分布 ($\bar{\lambda}_A e^{-\bar{\lambda}_A x}$) の差をとっている。これによって、映像全体を基準

として、 s_i 内の出現区間が相対的にどれだけ短くなっているか、もしくは長くなっているかを同時に評価している。最終的に、しきい値以上のバースト性をもつ s_i をトピックとして抽出する。

4 実験結果

本論文では、4 本の商用映画を実験映像として、主人公を対象としたインターバルシークエンスを作成した。まず、セグメンテーション結果に関しては、平均して約 77% のサブシークエンスに意味的なまとまりがあるという良好な結果が得られた。主な理由としては、登場人物と他の人物のインタラクションが特徴づけるために、出現区間と非出現区間の発生率が非常に有効に機能したことが挙げられる。具体的には、登場人物が単独のアクションをしている間は非出現区間はほとんどなく、他の人物が関係してくると出現区間と非出現区間が繰り返されるようになる。

別の理由としては、出現区間長と非出現区間長が、登場人物のアクションの変化にうまく対応していた点が挙げられる。例えば、登場人物が主体の会話、他の人物が主体の会話、盛り上がった会話という推移に応じて、出現区間長と非出現区間長が変化するという興味深い結果が得られている。上記の結果から、映像をインターバルシークエンスとしてモデル化することの有効性が分かる。

図 2 に、各映像から抽出されたトピックの例を示す。ここで、インターバルシークエンスは棒グラフの形式で表現されている。すなわち、各時区間を 90 度回転させて、出現区間を正の側、非出現区間を負の側に配置している。ゆえに、縦軸の正の値が出現区間長、負の値が非出現区間長を表わしている。また、縦の点線はサブシークエンスの境界、正の側にある横線は平均出現区間長、負の側にある横線は平均非出現区間長を表わしている。さらに、サブシークエンスごとのバースト性を、太線のパルス波の形式で示している。

図 2 から、出現区間長が異常に短く、もしくは長くなっているサブシークエンスがトピックとして抽出されていることが分かる。とりわけ、(a) の 11 から 13 番目、(c) の 14、15 番目のトピックではスリリングなムード、(b) の 20、21 番目のトピックではロマンティックなムードを演出するための編集技法が使用されている。また、(d) の 2、3 番目のトピックは、登場人物のおかしなアクション

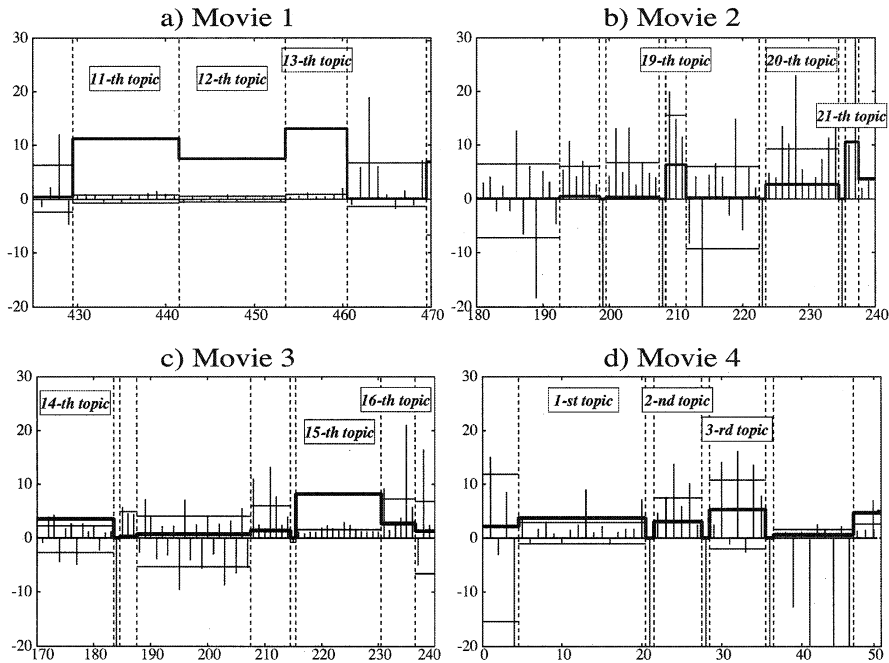


Fig. 2 抽出されたトピックの例

をじっくり映すことに対応して、出現区間長が異常に長くなるという興味深い結果を表わしている。このように、本手法を用いて、視聴者にインパクトを与えるトピックが多数抽出されていることが分かる。最終的に、抽出されたトピックは、映像インデキシング、映像ブラウジング、映像要約などに応用可能である。

5 おわりに

本論文では、フィクション映像を、登場人物の出現区間と非出現区間を表わすインターバルシーケンスとしてモデル化した。そして、出現区間長が異常に短くなる、もしくは長くなるパターンをバーストとして検出して、視聴者にインパクトを与えるトピックを抽出する手法を提案した。今後の課題としては、本手法を自動化するために、登場人物を高精度に認識可能な手法を開発することを計画している。

参考文献

- 1) Gemmell D. et al.: Multimedia Storage Servers: A Tutorial, *IEEE Computer*,

Vol. 28, No. 5, pp. 40 – 49 (1995).

- 2) Arijon D.: *Grammar of Film Language*, Focal Press Limited Publishers (1976). (岩本憲児, 出口丈人訳: 映画の文法, 紀伊国屋書店 (1980))
- 3) Kleinberg J.: Bursty and Hierarchical Structures in Streams, *Proc. of KDD 2002*, pp. 91 – 101 (2002).
- 4) Chatman S.: *Coming to Terms: The Rhetoric of Narrative in Fiction and Film*, Cornell University Press (1990). (田中秀人訳: 小説と映画の修辞学, 水声社 (1998))
- 5) Himberg J. et al.: Time Series Segmentation for Context Recognition in Mobile Devices, *Proc. of ICDM 2001*, pp. 203 – 210 (2001).