

動画シーンクラスタリングにおける 属性重み付け方法

3Q-7

秋元俊昭*, 岡夏樹**

*松下電器産業 東京通信システム研究所 **松下技研 情報・ネットワーク研究所

1. はじめに

最近、動画像データのデジタル化により、扱う動画像データの量も増加する傾向にあり、動画像データを効率よく分類し、検索する技術への期待が高まってきている。

こうした中、類似した事例は2次元マップの近傍に集まるという特徴を生かして、自己組織化マップ⁽¹⁾ (以下SOM法)を用いた動画像データの類似検索に関する研究⁽²⁾⁽³⁾も行われている。しかし、主眼はマップ表現や属性の工夫であり、属性の重み付けによって意味のある分類に近づけようとするものではない。

本研究の目的は、動画像シーンから抽出された複数の属性の重み付けを行い、SOM法により動画像シーンを分類する方法を提案することである。抽出された属性は、分類には全く役に立たない属性、概念によって重要度の異なる属性を含む。

本稿では、動画像シーン分類の適用例として、動画像シーンの類似検索を想定する。

2. 属性重み付け方法

あらかじめサンプルとなる動画像シーンを用意しておき、シーンの先頭画像あるいはシーンの音の特徴によって検索したい画像かそうでないかの区別を人が行う。このサンプルデータを訓練データとし、訓練データの分類がより正確になるような属性の重み付けを行い、未知のテストデータを分類する。ただし、検索したい画像は、例えばニュースのスタジオのシーンのように、意味のある画像の集合とする。

2.1 重み付け方法

サンプル動画像シーン集合の内、検索したい動画像シ

ーンの集合を $D_g = \{x_1, \dots, x_p\}$ 、そうでない動画像シーンの集合を $D_f = \{y_1, \dots, y_q\}$ とする。… (ステップ1)

次に、SOMを用いて重み付けされていない属性によってサンプル動画像シーンを分類する。… (ステップ2)

ステップ2の分類において、集合 D_g の事例と同じ分類になった集合 D_f の部分集合を集合 $E_f \subset D_f$ とする。

… (ステップ3)

時刻の初期値を $t=0$ とする。… (ステップ4)

集合 D_g の事例 x_i の属性ベクトルを $v_{x_i} = (r_{1x_i}, \dots, r_{N_{x_i}})$ とし、集合 E_f の事例 y_i の属性ベクトルを $v_{y_i} = (r_{1y_i}, \dots, r_{N_{y_i}})$ とする。

ある時刻 t に、集合 D_g から異なる2つのシーンを選択し、選択したシーンの属性の差分2乗ベクトルを $A_j(t) = ((r_{1x_k} - r_{1x_i})^2, \dots, (r_{N_{x_k}} - r_{N_{x_i}})^2) (x_k \neq x_i)$ とする。また、ある時刻 t に集合 D_g から1つ、集合 E_f から1つシーンを選択し、選択した2つのシーンの属性の差分2乗ベクトルを $B_m(t) = ((r_{1x_n} - r_{1y_o})^2, \dots, (r_{N_{x_n}} - r_{N_{y_o}})^2)$ とする。時刻 t における属性重みベクトルを $W(t) = (w_1(t), \dots, w_N(t))$ とすると

$$W(t+1) = W(t) \times \left(1 - \beta \times \frac{A_j(t) - B_m(t)}{A_j(t) + B_m(t)} \right) \dots \text{(式1)}$$

式1により、重みを更新する。… (ステップ5)

時刻 t がある値以下であれば、 $t=t+1$ として、ステップ5に戻る。そうでなければ、動作を停止する。

… (ステップ6)

2.2 属性の重み付け方法の説明

ステップ1では、類似検索用の動画像データを選択する。ステップ2、3では、集合 D_f の事例 x_i に類似した、集合 E_f の事例 y_i に限定することにより、学習効率が向上することが予備実験で確かめられた。ステップ4の組み合わせは、全ての組み合わせが平均化されるようにする。ステップ5では、属性値はあらかじめ[0,1]で規格化して

おく。 $A_j(i) - B_m(i)$ は、この属性において、クラス内の類似度と、クラス間の類似度の比較に相当する。 β は学習係数であり、経験から得られた値として0.001程度に設定する。ステップ6で設定する繰り返し回数の目安として、全ての組み合わせに相当する回数を設定する。

3. 実験、評価

3本のニュース映像を用い、2本を訓練データ、1本をテストデータとして3つのテストセットにより実験を行った。

シーン検出ソフト⁽⁴⁾により、番組1本当たり約200のシーンが検出され、シーンの先頭画像から64色のヒストグラム（64属性）、25分割画面のエッジ数、全画面のエッジ数（26属性）、あるいはシーンを構成する音(MPEG Audio level2)から、ケフレンシー⁽⁵⁾（分析窓長256msec、分析周期100msec）の最上位の値を5段階に量子化した値の割合（5属性）、全部で95の属性値を抽出した。

類似検索用の動画データとして、スタジオシーンと効果音によるテーマ切り替わりシーンの2つを用いた。

重みの評価方法としては、類似検索による適合率(P)、再現率(R)（両者を重み評価値とする）を測定し、3つのテストセットの平均を計算した。また、同じ属性重みを用いて、SOM法とk-NN法によって類似検索を行い、評価を行った。

SOMによる類似検索の場合、4x3のマップにおいて、訓練データの検索シーンが最も多く出力された出力ユニットと同一の出力ユニットを検索範囲とし（マップのサイズが小さいことを考慮）、k-NN法では、ノイズデータを配慮してk=3とし、訓練データを記憶データとして実験した。

効果音によるテーマ切り替わりシーンをういた重み付けの結果を図1に示す（重みの初期値は0.5）。また、スタジオシーンをういた重み付け学習回数と3-NN法による重み評価値、SOM法による重み評価値の関係をそれぞれ図2(a)、(b)に示す。

学習回数と属性重みの関係は、学習回数を重ねるにつれ、分類に役立つ属性値重みは初期値より増加していき、

逆に分類に役立たない属性値重みは初期値と同じか減少していく。

SOM法に比べ、k-NN法の方が適合率は良く、繰り返し回数に対する重みの評価値の変化も安定している。これは、分類のための空間分割がk-NN法の方が細かいからであると思われる。

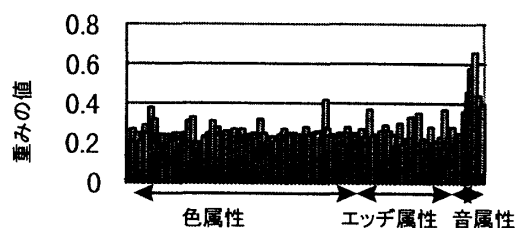
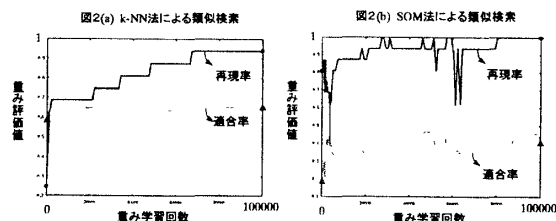


図1: 効果音によるテーマ切り替わりシーンの重み分布



4. むすび

今回の属性の重み付け方法によって、分類精度は向上したが、分類は2つのクラスに限定されており、必ずしもSOM法に適した重み付けとは言えない。今後は、3つ以上のクラスへの対応や、参照ベクトルの決定方法を考慮した重み付け方法の検討を行っていきたい。

謝辞 九州工業大学石川真澄教授との議論が本研究のきっかけになりましたので、お礼申し上げます。なお本研究は一部IPAの「創造的ソフトウェア育成事業」による支援を受けています。

参考文献

- (1) T. Kohonen 著、徳高ら訳、「自己組織化マップ」、シュプリンガー・フェアラーク東京
- (2) 波多野、田中、「映像データベースの動的クラスタリングと素材検索機構について」、情報処理学会データベースシステム109-18 p.105-110 1996
- (3) Kyung-Ah, "Image Organizing and Retrieval with Automatically Constructed Feature Vectors", ACM SIGIR96, August 18-22, 1996 Zurich Switzerland
- (4) 有田、山田、鈴木、「マルチメディアオフィスシステムMUSCUT」、National Technical Report, Vol. 42, No. 5 Oct. 1996
- (5) 古井貞熙、「音響・音声工学」、近代科学社