

Word2Vec を用いて歌詞と写真を対応づけた スライドショー生成システム

梅村 允康[†] 保利 武志[‡] 嵯峨山 茂樹^{†‡}

[†] 明治大学 総合数理学部 [‡] 明治大学 大学院先端数理科学研究科

1 はじめに

近年, YouTube などの動画共有サービスや SNS 利用者の増加により, 一般ユーザーが写真や動画を撮影, 編集し映像作品を作成する機会が増えている. その際, 写真や動画に対し音楽を付与することで, 視覚と聴覚を刺激しさらに印象的かつ魅力的な映像作品ができると思われる. また一般的に映像作品の中でも画像同士を繋げたスライドショーは芸術的にも実用的にも利用され, 比較的作成が容易であるという利点がある.

歌詞と画像を関連づけたスライドショー作成に関する研究は, これまでに歌詞に出現する名詞と歌詞全体を表現する印象語を基に画像を決定する手法が提案されてきた [1]. 本研究では, 歌詞に合った画像を選定することによる, より印象深い映像作品の実現を目的とし, ユーザーが指定した楽曲と画像に対し, Word2Vec を用いて楽曲の歌詞と画像を対応付けたスライドショー生成システムを提案する.

2 スライドショー生成

2.1 問題の解析

「歌詞と画像の内容が適合していて, 且つ近い印象の画像に遷移することでより滑らかで自然なスライドショーが完成する」と仮定すると, 必要なのは歌詞と画像のマッチングと, 画像印象の数値化及び距離尺度を定式化することである.

2.2 歌詞と画像の対応付け手法

画像の内容と歌詞を関連づけるためユーザーが画像の印象を 5 つのラベルとして付与し, そのラベルと歌詞情報 DB に対し Word2Vec を用いることで近い印象を持つ語同士が近傍に配置されるような歌詞/ラベル意味空

間を構築できる (単語ベクトル DB). 単語ベクトル DB に基づき, 拍ごとに画像のラベルと歌詞の \cos 距離を求め正規化することで, ある画像においてある歌詞が出力される確率 $f_k(x)$ (k : 拍数, x : 画像) が求まる. ここで, スライドショーを作成するにあたって, 画像を切り替えるタイミングや枚数を予め設定しておく必要があるため, ユーザーは指定した楽曲の拍を取得し, 切り替えるタイミングを設定する.

2.3 画像印象の数値化及び画像の遷移

画像間の類似度を, 以下の 2 手法で求め, 正規化した結果を遷移確率 $p(y|x)$ (画像 x から画像 y に遷移する確率) として定めた.

I ヒストグラム

画像同士を比較するために相関, カイ 2 乗値, 及び Bhattacharyya 距離をそれぞれ求め正規化し, 3 つの平均を求めることで類似度を求めた.

II 形状マッチング

輝度値の変化に強い形状情報として, 形状の比較に Hu モーメント不変量を用いた. これは回転・スケール・反転に対して不変という利点がある.

2.4 画像の選定方法

2.2 と 2.3 で求めた確率に対し, 拍ごとの部分問題と捉えると, 動的計画法を用いて効率的に尤度最大となる画像の組を得ることができるが, その際に同じ画像が複数回選ばれてしまう可能性がある. この問題は巡回セールスマン問題 (traveling salesman problem; TSP) と等価として捉えることができるため, TSP の近似解を求める手法として代表的な遺伝的アルゴリズム (genetic algorithm; GA) を用いて画像の順序を決定する. 一般的な GA の手順に従い, 具体的には以下に行った [2].

I 初期集団の生成

画像の枚数を n 枚 ($n \geq$ 拍数) とし, 画像 1 から画像 n のうち拍数分を拍ごとに割り当てたものを巡回路とした. また初期集団としてランダムな巡回路を 30 回路用意した.

“Correspondence of lyrics and photos slide show generation system using Word2Vec”

by Nobuyasu Umemura[†], Takeshi Hori[‡] and Shigeki Sagayama^{†‡},

[†]School of Interdisciplinary Mathematical Sciences, Meiji University, [‡]Graduate School of Advanced Mathematical Sciences, Meiji University.

II 評価

巡回路の1拍目の画像に遷移する確率(初期遷移確率)を $p(x_1)$, 評価値を V とすると,

$$V = f_1(x_1)p(x_1) \prod_{i=2}^k f_i(x_i)p(x_i|x_{i-1}) \quad (1)$$

と定式化でき, V が大きいほど優れた巡回路と見なすことができる。

III 選択

ある巡回路 i の評価値を V_i とすると,その巡回路が選択される確率 $S(i)$ (J :回路数)を以下のように与えた:

$$S(i) = \frac{V_i}{\sum_{j=1}^J V_j} \quad (2)$$

IV 交叉

同じ画像が複数回選ばれる致死個体を避けるため,順序交叉を用いた。

V 突然変異

10%の確率で巡回路の中でランダムに選ばれた2つの画像を入れ替えた。

3 評価実験

3.1 実験条件

歌詞と画像を対応づけることにより生成されたスライドショーを評価するため,歌詞と画像との対応付けがあるものとなないものとの比較実験を行った。実験協力者は, J-POP 楽曲1曲に対して,歌詞と画像との対応付けがあるものとなないものそれぞれ2つのスライドショーについて,以下の3項目に対して10段階 MOS 評価に基づく主観評価実験を行った。実験には10代から20代の男女10名が参加した。

- Q1. スライドショー自体の完成度。
- Q2. 歌詞と画像が合っているか否か。
- Q3. 画像の遷移が滑らかか否か。

3.2 実験結果

実験結果を図1, 2, 3に示す。本結果から, Word2Vec を用いて歌詞と画像を対応づけた方が主観評価が全般的に高いことが明らかになった。特に, Q2(歌詞と画像の適合: 図2)に関しては大きな差が見られた。Q1(スライドショーの完成度: 図1)については必ずしも有意な差は見られなかった。

4 まとめと課題

本研究では, 提案手法を用いて評価実験を行った結果, 歌詞の意味を考慮したスライドショーを作成可能である

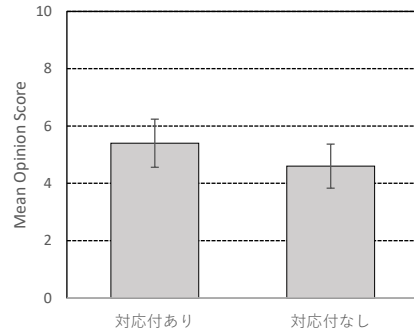


図1 Q1(スライドショーの完成度)の主観評価結果比較

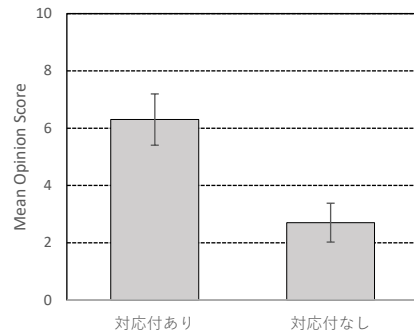


図2 Q2(歌詞と画像の合致)の主観評価結果比較

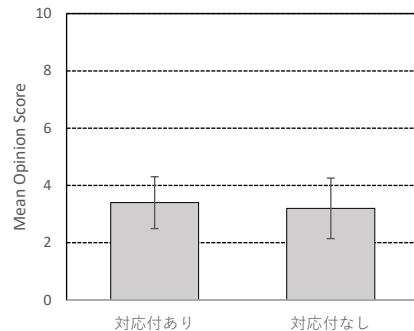


図3 Q3(画像遷移の滑らかさ)の主観評価結果比較

ことが示された。今後の課題として, 画像の枚数や画像のラベル数を増やすことや, 本稿で有意性を示すことができなかった画像の類似度の再検討などを行いたい。またトランジションや表示方法など, スライドショー自体の改善も行う予定である。

謝辞 本研究は JSPS 科研費 17H00749 の助成を受けた。

参考文献

- [1] 船澤 *et al.*, “歌詞情報を利用した web 画像・楽曲連動スライドショー自動生成システム,” 研究報告音楽情報科学 (MUS), pp. 1–6, 2010.
- [2] J. J. Grefenstette *et al.*, “Genetic algorithms for the traveling salesman problem,” *Proceedings of the 1st International Conference on Genetic Algorithms*, pp. 160–168, 1985.