

大規模歌詞データからの潜在的トピック遷移のモデル化

渡邊研斗[†] 松林優一郎[†] 乾健太郎[†] 後藤真孝[‡]
 東北大学[†] 産業技術総合研究所[‡]

{kento.w, y-matsu, inui}@ecei.tohoku.ac.jp m.goto@aist.go.jp

1 はじめに

近年コンピュータを使用した音楽制作技術の発達に伴い、プロの作曲家でなくとも音楽を創作できる環境ができ、今後は歌詞や楽曲の自動創作・創作支援技術の発展が期待されている。しかし歌詞生成に関する関連研究は少なく、韻や拍子などのスタイルに合わせた歌詞を生成する研究が殆どである [1, 2, 3]。本稿では歌詞の中に潜在的に存在する典型的なトピックの遷移パターンを、大規模な歌詞データから統計的に教師なし学習するモデルを提案し、その振る舞いを分析する。また歌詞作成支援への応用方法についても論じる。

2 歌詞内容の展開 (トピック遷移) の学習

作詞家は「Aメロ Bメロ サビ」に対して「話の場面設定 回想 主張・感情」のような内容の展開を意識することで、共感しやすい歌詞を制作している [4, 5, 6]。本研究では、歌詞の内容の展開 (以後トピック遷移と呼ぶ) をコンテンツモデル [7] を応用してモデル化する。

2.1 コンテンツモデル

コンテンツモデルは隠れマルコフモデルの拡張型であり、複数のトピックが混在する文章の要約や分割に用いられている [7, 8]。図1に本研究で設計したコンテンツモデルの簡略図を示した。本研究では各隠れ状態 (トピック) から、歌詞データ内で予め分割されている段落を生成するモデルを設計した。

ここで変数 $t = 1, 2, \dots, T$ は入力歌詞での段落の位置を表しており、変数 $n = 1, 2, \dots, N$ は各段落内の単語の位置を表している。予め段落ごとに分割されている歌詞 $Lyrics = Paragraph_1, \dots, Paragraph_t, \dots, Paragraph_T$ に対してトピック遷移列を $\mathbf{z} = z_1, \dots, z_t, \dots, z_T$ とする時、位置 t のトピック z_t は以下に示す遷移確率分布 $\pi_{z_{t-1}}$ によって生成される。

$$z_t \sim \pi_{z_{t-1}} = \frac{\text{count}(z_t, z_{t-1}) + \alpha}{\sum_{k=1}^K \text{count}(z_k, z_{t-1}) + K * \alpha} \quad (1)$$

ここで K は学習前に予め決めておくトピックの数であり $\text{count}(z_i, z_j)$ は z_j から z_i への遷移回数を表す。また、この遷移確率はハイパーパラメータ α によって平滑化されている。各位置のトピック z_t から段落 $Paragraph_t$ の単語 w_n を生成する確率分布 θ_{z_t} は以下の式で表される。

$$w_n \sim \theta_{z_t} = \frac{\text{count}(w_n, z_t) + \beta}{\sum_{w \in W} \text{count}(w, z_t) + W * \beta} \quad (2)$$

ここで W は学習時に使用する語彙であり、 $\text{count}(w_n, z_t)$ は位置 t におけるトピック z_t において単語 w_n が生成された回数である。また、単語生成確率はハイパーパラ

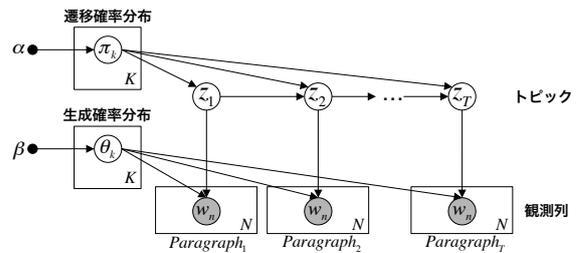


図1: コンテンツモデルの簡略図

メータ β によって平滑化されている。以上の遷移確率と生成確率を用いることにより、歌詞全体を生成する確率 $P(Lyrics)$ は以下の式で定義される。

$$P(Lyrics) = \sum_{\mathbf{z}_0^T} \prod_{t=1}^T \pi_{z_{t-1}} \prod_{n=1}^N \theta_{z_t} \quad (3)$$

ここで \mathbf{z}_0^T は起こりうる全てのトピック遷移系列である。そのため、歌詞生成確率 $P(Lyrics)$ を計算するために、本研究では動的計画法に基づいた前向きアルゴリズムを用いて歌詞生成確率を求めた。

本研究では、局所解に陥らなく性能が良いモデルを学習ができることが知られているギブスサンプリングを採用した [9]。トピック遷移の推定には、ピタビアルゴリズムと呼ばれる動的計画法を用いることで、最も生成確率が大きいトピック遷移列を効率的に推定できる。

3 評価・分析

コンテンツモデルがトピック遷移を学習できているかを、以下の手順で定量的に評価した。

1. テストデータの各歌詞について、すべての可能な文節の順列を列挙する。
2. 各順番について、式 (3) を用いて生成確率を計算する。
3. 最も生成確率が高い順番を、モデルが最も適切と判断した順番 A とする。
4. 適切と判断した順番 A とテストデータでの順番 (正解) B を比較し、下記の式 (4) により τ を計算する。

$$\tau = \frac{\text{count}^+(A, B) - \text{count}^-(A, B)}{T(T-1)/2} \quad (4)$$

ここで $\text{count}^+(A, B)$ は順番 A から抜き出した二つの文節ペアと、順番 B から抜き出したペアが同じになる回数であり、 $\text{count}^-(A, B)$ は異なる回数である。 τ は Kendall の順位相関係数と呼ばれ、-1 から 1 までの値を取り、1 に近づくほど二つの順番に相関があり -1 に近づくほど逆の相関があることを意味する。本研究では、歌詞内の段落を適切に並び替えられるようなモデルがトピック遷移を捉えていると考えたため、この指標を採用した。

表 1: 段落間のトピック遷移確率 (上位 3 つの遷移確率を太字で示す)

		遷移先のトピック										
		END	啓発	恋愛	演歌	口語	願望	情景	ポジティブ	リズム	英語 1	英語 2
遷移元のトピック	START	-	8.36	4.55	17.00	13.47	6.05	37.36	3.93	5.26	2.58	1.43
	啓発	10.20	47.39	6.84	0.28	3.46	6.39	11.32	9.21	2.33	1.49	1.09
	恋愛	18.36	7.79	39.03	0.02	2.85	4.50	13.44	2.83	2.98	6.14	2.04
	演歌	28.68	0.28	0.00	66.42	1.95	0.47	1.46	0.60	0.05	0.00	0.03
	口語	10.51	4.70	2.45	1.85	61.56	3.87	6.80	0.73	5.58	0.76	1.18
	願望	16.47	6.64	5.14	0.46	3.05	42.20	20.11	3.58	0.33	1.13	0.86
	情景	9.01	8.72	6.90	1.06	3.79	12.18	47.94	5.93	1.99	1.08	1.40
	ポジティブ	20.69	13.04	4.57	0.45	0.60	5.33	14.38	37.16	1.04	1.70	0.97
	リズム	13.96	4.54	5.20	0.00	10.23	0.42	6.03	1.20	50.45	3.55	4.40
	英語 1	22.23	4.49	8.48	0.00	3.04	2.15	7.56	2.48	5.87	43.08	0.51
英語 2	23.90	5.32	5.83	0.13	6.31	3.11	12.14	1.60	9.90	1.25	30.29	

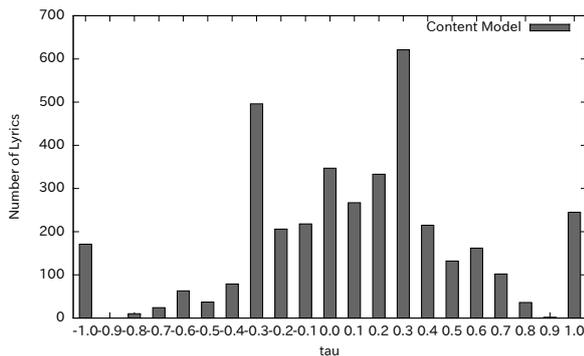


図 2: ケンドールの順位相関係数のヒストグラム

3.1 パラメータチューニング

本研究では日本のポピュラー音楽 26,196 曲の歌詞を使用し、学習用の訓練データ・パラメータチューニング用の開発データ・評価用のテストデータを 6:2:2 の割合で分割した。モデルの学習ではギブスサンプリングの繰り返し数を 1000 回とし、平滑化のためのハイパーパラメータ α, β を 0.01 とした。学習時は歌詞の内容に関係しているであろう「名詞」「動詞」「形容詞」「副詞」「助動詞」の単語のみで学習した。また、トピックの数 K は開発データでの評価性能が高かった $K = 10$ を採用した。

3.2 評価結果

図 2 にテストデータの τ のヒストグラムを示した。この図より、 τ が正の値である歌詞の数が多いことがわかる。また、 τ の平均は 0.103 となった。この値は決して高い正の相関ではないが、モデルがトピック遷移を捉えている傾向が見られた。

3.3 モデル分析

どのようなトピック遷移が学習されたか分析するために、表 1 に各トピックの遷移確率を示し、表 2 に各トピックの代表的な単語を示した。まず表 2 から「情景」「願望」など、意味にまとまりのあるトピックが学習されたことがわかる。また、表 1 より以下の特徴が見られた。

- どのトピックも自己遷移確率と曲の終了への遷移確率が大きい。
- 演歌トピックは他のトピックには遷移しにくい。
- 情景トピックは他のトピックから遷移されやすい。

一見、トピック遷移が無いように見られるが、「情景」トピックへの遷移には明らかな偏りがあることがわかる。これは「サビ」の内容を共感させやすくするための「場面設定」や「回想」の役割を、「情景」トピックが果たしているためだと考えられる。

表 2: 各トピックの代表的な単語

トピック	代表的な単語
啓発	自分, 手, 生き, 道, 見え, 目, 前, 幸せ, 日々, 全て
恋愛	love, believe, rain, boy, 止め, girl, cry, 止まら
演歌	女, 恋, 花, 涙, 男, 人, ます, 夜, 酒, ぬ, です, 雨
口語	た, な, いい, たら, する, なっ, 今日, だっ, どう
願望	たい, 今, ずっと, 心, 愛し, 好き, 忘れ, 抱きしめ
情景	街, 見, 夏, 瞳, 消え, 時間, 月, 窓, 白い, 思い出
ポジティブ	愛, 夢, 空, 風, 星, 胸, 明日, 光, 未来, 信じ, 世界
リズム	yeah, hey, oh, wow, yes, さあ, ほら, 行く, リズム
英語 1	la, love, can, now, baby, 'll, go, heart, know
英語 2	wow, go, oh, love, go, me, let, my, do, come, get

4 歌詞生成支援への応用について

本研究で設計したモデルは、任意の歌詞に対してトピック遷移列を推定することができるため、歌詞の流れを考慮した作詞支援が期待できる。例えば「Bメロ」だけ出来ていない未完成の歌詞を補充したり、既存の歌詞と似た展開の歌詞を生成できることが考えられる。

5 おわりに

本研究では、教師なし機械学習法を用いて 2 万曲以上の歌詞からトピック遷移を学習した。評価・分析の結果から、モデルが「情景的」なトピックと「願望」「恋愛」などの「感情的」なトピックとの間の遷移を学習していることがわかった。

参考文献

- [1] Erica Greene, Tugba Bodrumlu, and Kevin Knight. Automatic analysis of rhythmic poetry with applications to generation and translation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 524–533, 2010.
- [2] Ananth Ramakrishnan A and Sobha Lalitha Devi. An alternate approach towards meaningful lyric generation in tamil. In *Proceedings of the NAACL HLT 2010 Second Workshop on Computational Approaches to Linguistic Creativity*, pp. 31–39, 2010.
- [3] Gabriele Barbieri, François Pachet, Pierre Roy, and Mirko Degli Esposti. Markov constraints for generating lyrics with style. In *Proceedings of the 20th European Conference on Artificial Intelligence*, pp. 115–120, 2012.
- [4] 森浩美, 宇多丸, ヒロイズム, MiChi, zopp, みろく, 吉本由美, いしわり 淳治. 作詞のための 8 の極意. Yamaha Music Media, 2014.
- [5] 田口俊. 思いどおりに作詞ができる本. リットーミュージック, 2012.
- [6] 上田起士. よくわかる作詞の教科書. Yamaha Music Media, 2010.
- [7] Regina Barzilay and Lillian Lee. Catching the drift: Probabilistic content models, with applications to generation and summarization. In *Proceedings of the 2004 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 113–120, 2004.
- [8] 但馬康宏. 言語モデルの違いによる hmm を用いたテキストセグメンテーションの性能比較. 情報処理学会論文誌, pp. 38–46, 2013.
- [9] Alan Ritter, Colin Cherry, and Bill Dolan. Unsupervised modeling of twitter conversations. In *Proceedings of the 2010 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 172–180, 2010.