

単語どうしの共起関係を用いた 物語テキストの自動分割手法とその評価 - 歌詞の自動生成を目指して -

Automatic segmentation method and evaluation based on co-occurrence relation of words in a story text

山口 葵 竹川 佳成 平田 圭二
Aoi Yamaguchi Yoshinari Takegawa Keiji Hirata

公立はこだて未来大学
Future University Hakodate

The purpose of this study is to automatize the division of folk tales and short-short stories into the parts of a well-organized standard story structure. We use the set of the short-short stories written by Shinichi Hoshi as a story corpus. We develop a few algorithms to automatically divide their texts into several scenes. The algorithms work referring to the result of preprocessing by human and the co-occurrence relationship of words. In this paper we discuss the characteristics of each algorithm, applying to the corpus.

1. はじめに

近年, DTM (Desktop Music) の普及により, 楽器を所持せずとも手軽に楽曲制作のできる環境が整ってきた. 例えばニコニコ動画などの動画投稿サイトには, ボーカロイド^{*1}を用いたオリジナル楽曲のPVが多数アップロードされている. しかし環境はまだ不十分であり, 初心者がゼロから楽曲を完成させることは容易ではない. 特に声楽曲を制作したい場合は作曲に加え作詞も考える必要があるが, 専門的知識や技術の無い者にとっては, どちらも十分に質の高い作品を作ることは難しい. そこで本研究では特に作詞に着目し, 歌詞と成り得る文章を自動生成する仕組みを考える. まず既存の歌詞に見られる, 機械によって実現可能性のある特徴について考える. 歌詞には多種多様な形があるが, 基本的に

- 全体の文字数が一般的な小説などの文字数より少ない
- 一貫したテーマや起承転結を意識した構造を持つ

という点は共通していると考えられる. この2点が顕著に表れている例として, 「桃太郎」や「うさぎとかめ」といった昔話や童話を題材にした童謡が挙げられる. これらの歌詞は物語の内容の流れに沿って短くまとめたものとなっており, 起承転結の構造を持っている. また, 元となった物語よりも文字数が圧倒的に少ない場合がほとんどである. つまり物語を元に作成された歌詞は, その物語内容を凝縮した要約のようなものとも言える. よって, 従来の文書要約技術を適用できる可能性がある.

上記を踏まえ, 物語テキストを入力とすることで, その内容を元にした歌詞を自動生成するシステムを提案する. 本稿ではその一部分である, 物語テキストを起承転結に自動で分割する処理について主に述べる. なお物語テキストとは, 例えば童話や星新一のショートショートなど, 起承転結を持つ物語を指す. 入手の容易さから, コーパスとして星新一のショートショートを用いる.

連絡先: 山口 葵, 公立はこだて未来大学, b1011168@gmail.com

*1 ヤマハが開発した歌声合成技術および, その応用ソフトウェア (<http://www.vocaloid.com/>)

2. 関連研究

テキスト分割に関する先行研究について述べる. 小林は物語から場所・時間・登場人物を含む文を抽出し, それらの重なり具合からシーン境界の推定を行っている [1]. ただし準備の段階で, ある語句が場所や時間の概念を含むかどうかを判別しておく必要がある.

また赤石は単語同士の共起関係から算出される吸引力を元にトピックの遷移を推定し, 文書の複数シーンへの分割を試みている [2]. ある語 t の吸引力 $\text{attr}(t)$ とは, 語 t が同じ文書中の他の語を引き付ける力を表す. ある語 t の吸引力を求めるには, まず語 t の t' に対する出現依存度, $\text{td}(t, t')$ を求める. これは, 語 t と同じ文中に語 t' が出現する条件付き確率である. ただし t, t', T は, 文書に含まれる全単語の集合とする.

$$\text{td}(t, t') = \frac{\text{sentences}(t) \cdot \text{sentences}(t')}{\text{sentences}(t)}$$

ここで $\text{sentences}(t)$ は文書中において語 t を含む文の数を, $\text{sentences}(t) \cdot \text{sentences}(t')$ は語 t と語 t' を同時に含む文の数を示す. 語 t と, 文書中に出現する t 以外の全ての語との出現依存度の総和が, 吸引力の大きさとなる.

$$\text{attr}(t) = \sum_{t'' \in T} \text{td}(t'', t)$$

対象文書の例として, ギリシャ神話や「羅生門 (芥川龍之介)」が取り上げられている. しかし本研究では, それらよりも圧倒的に文字数の少ない星新一ショートショートを取り扱うため, 表れるトピック数も少なくなると考えられる. そこで人間がテキスト分割を行う際の考え方を取り入れることで, 文字数の異なるテキストにも対応可能な処理を考える.

3. 処理方式

図1に, 入力した物語テキストファイルから歌詞を得る流れを示す.

3.1 名詞の抽出

物語テキスト中の登場人物や固有名詞といった単語は, その物語を特徴付けていると考えて良いであろう. そこで形態素解

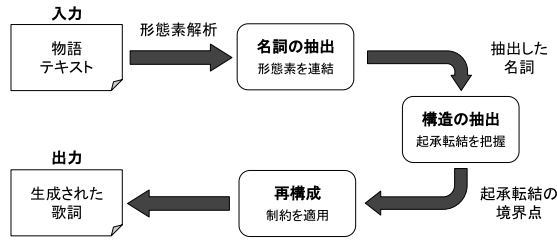


図 1: 歌詞自動生成における処理

析器 lucene-gosen^{*2} を用いて、テキスト中から名詞を抽出する。しかし名詞によっては、形態素への分割が意図した以上に細かくなってしまふ場合がある。例えば名詞「神経科医」を形態素解析すると、神経・科・医の3つの形態素に分割される。これでは名詞の意味が損なわれてしまうと考えられる。このような事態を防ぐため、品詞情報を元に形態素の連結を行い1つの名詞とした。連結の条件は、森らを元に以下のとおり設定した [3]。

- 「名詞」とは、「名詞または人名」かつ「非自立・代名詞・ナイ形容詞語幹・助動詞語幹・接頭・接尾・特殊」を含まないものとする
- 「名詞」が連続する場合は連結する
- 「名詞」の後に「接尾」が続く場合は連結する
- 「名詞」の後に「助数詞」が続く場合は連結する
- 「数字」または「アルファベット」が続く場合は連結する

3.2 構造の抽出

物語内容を踏まえた歌詞を生成するため、元となる物語テキストの構造を踏襲する。歌詞の起に当たるパートを作成する時は、同じく物語テキストの起に当たるパートで出現する単語を主に利用する。簡易のため、今回は入力として与える物語テキストを、起承転結の構造を持つものに限定する。物語テキストを起承転結の4シーンに分割するために、単語同士の共起関係から算出される吸引力(2章)を用いる。

文書を何文目まで読み込んだかによって、吸引力の値は変化する。例えば表1は、星新一ショートショート「こん」の1~3文目において、名詞がそれぞれ何文目に出現するかをまとめたものである。この例で名詞「男」の吸引力を計算する場合を考える。まずテキストを1文目まで読み込んだ時は、「男」という名詞がまだ出現していないため吸引力の値は0である。次に2文目まで読み込んだ時は、1~2文目までに出現する「男」以外の全名詞、すなわち「つぎ、神経科医、声、診察室、女」の5語との出現依存度を計算し積分する。この時の吸引力の値は4となる。また3文目まで読み込んだ時は、1~3文目までに出現する「男」以外の全名詞「つぎ、神経科医、声、診察室、女、目、顔つき」の7語を用いることとなり、この時の吸引力の値は6となる。

このような文ごとの吸引力の変化を見て、上昇部(吸引力が増加している部分)から停滞部(吸引力が変化しない、あるいは減少する部分)へと変化する箇所を、文書のトピックが着目語から別の単語へと移行したと見なし、文書の分割箇所として検出する。これは赤石の手法 [2] に従った分割基準である。

*2 <https://code.google.com/p/lucene-gosen/>

表 1: 星新一「こん」1~3文目に出現する名詞

	1 文目	2 文目	3 文目
つぎ		x	x
神経科医	x		x
声	x		x
診察室	x		x
男	x		
女	x		
目	x	x	
顔つき	x	x	

34 酒に酔ったせいかと思ったが、そうでもないらしい。
 35 ミーラ星人たちは老人に「それをゆずってくれ」と、
 手まねでたのんだ。
 36 だが、老人は首と手を振った。

図 2: 星新一「ボウシ」より一部抜粋

3.3 再構成

抽出した名詞や構造といった物語テキストの特徴、及び歌詞としての制約を考慮し、歌詞と成り得る文章を構成する。また歌詞らしさを付与するための制約を一般化し、適用する。歌詞は、言葉を発音した時のリズムが良かったり、文章が韻を踏んでいるなど、メロディに乗せた時に違和感なく歌うことのできるものが好ましい。例えば、物語のある1文をそのまま抜き出して適当なメロディを付けたとしても、元は歌詞として作られた文章ではないため、あまり歌詞らしさを感じることができないと考えられる。なお、この処理は本研究では未実装である。

4. シーン分割アルゴリズムの実験

物語テキストを起承転結の4つに分割するためのシーン分割アルゴリズムを実装し、実際の物語テキストに対し分割実験を行った。今回は

手法1: 吸引力(2章)による分割

手法2: 対象範囲を狭めた吸引力による分割

手法3: 正解データの情報を援用した分割(手法3-1, 3-2)

の3つの手法を試した。いずれも、承・転・結の各シーンの始まる可能性のある文の番号(以降、これを分割点と呼ぶ)を出力とする。なお、起の開始点はどの作品でも物語テキストの最初の1文目で固定されているため、出力には含めない。

4.1 吸引力による分割

この分割手法では、まず入力テキスト中に出現する全ての名詞に対し、1文目~テキスト最終文まで読み込んだ時の吸引力を算出した。その後、全名詞の吸引力の推移に着目し分割を行った。n文目の吸引力の大きさに着目した時、

$$\begin{aligned} (n-1 \text{ 文目の吸引力} < n \text{ 文目の吸引力}) \\ (n \text{ 文目の吸引力} > n+1 \text{ 文目の吸引力}) \end{aligned} \quad (1)$$

であれば、n+1文目を新しいシーンの開始点と見なす。他に、

一旦は増加した吸引力が変化しない状態を続けた後に減少するパターンがある。この場合も全体で見れば増加 減少という変化を辿っていると見なし、同じように新しいシーンの開始点と判断する。

$$\begin{aligned} (n-1 \text{ 文目の吸引力} = n \text{ 文目の吸引力}) \\ (n \text{ 文目の吸引力} > n+1 \text{ 文目の吸引力}) \end{aligned} \quad (2)$$

しかし 1 つの名詞のみに着目した場合、その名詞の出現頻度や位置などに影響を受ける可能性もある。それを考慮し、特定の名詞の吸引力の推移だけでなく、ある文まで読み込んだ時の全名詞の吸引力の合計値を算出し、その推移にも着目して同じ基準で分割を行った。

4.2 対象範囲を狭めた吸引力による分割

前節の手法では吸引力を算出する際、読み込んだ文までに出現する全ての名詞との共起度を利用していた。しかし人間がテキストを分割しようとする時、ある文がシーンの切れ目かどうかを考えるにあたり 1 文目から毎回読み直すようなことはしない。恐らく、着目文の前後のいくつかの文の内容を手がかりにして判断するであろう。そこで、この考え方をアルゴリズムに反映させるため、シーン分割点の前後の情報のみを利用する分割を行った。この手法ではある文に出現する名詞の吸引力を計算する場合、その文を中心に前後の一定数の文のみを参照し、その区間に出現する名詞を用いて吸引力を計算する。

この区間の大きさ（計算のために参照する文の数）を、本章では幅と表現する。今回はこの幅を、入力テキストの総文数 ÷ 4 に統一した。これは、1 つのテキストを 4 つのシーン（起承転結）に分割することを目指しているためである。この分割においても前節と同様、1 つの単語だけでなく全名詞の吸引力の合計値の推移にも着目している。ある文を分割点として検出する基準についても変更は無い。

4.3 正解データの作成

3 つ目の分割手法について述べる前に、使用した正解データについて述べる。ここで言う正解データとは、人間の手により起承転結に分割された物語テキストのことを指す。狙いは人間的な視点を取り入れることで、テキストの内容に沿ったより自然なシーン分割をすることである。また吸引力だけでは分割点である箇所は検出できても、それらが承・転・結のどの分割点に該当するかまでは判別できない。これも正解データの統計情報を用いることで、吸引力により分割点と判断した箇所がどのシーンの開始点として妥当かを判別することができる。

正解データの作成手法について記述する。まず分析者である 4 名の大学生及び大学院生に、星新一ショートショートの記事 20 篇を複数のシーンに分割してもらった。この時の分割の基準として村井らの設定した以下の 5 つの条件を用いた [4]。

- 本文での空行による分割（著者自身による物語の分割）
- 物理的な場面移動（会話内や想起されたエピソード内での移動は除く）
- 時間経過（短時間の物は除く）
- 主要な登場人物の出現・退出・移動、誕生・死去
- 状況説明の終了（短い説明は除く）

シーン数はいくつになっても構わないものとした。そして分割終了後に、この基準により付けられた分割点の中から、テキスト全体を見た際に承・転・結（起はテキスト 1 文目から開

始と見なすので除外）の始まると考えられる箇所を選んでもらった。

この起承転結の 4 シーンに分割されたテキスト（20 篇 × 4 人分 = 80 篇）それぞれに対し、各シーンがテキスト全体において何%を占めているかを算出した。4 人分の平均を取ったところ、テキスト全体で起の部分は 18%、承の部分は 35%、転の部分は 32%、結の部分は 15% となった（図 3 のグラフを参照）。また標準偏差は、起の部分は 6.9%、承の部分は 9.6%、転の部分は 8.2%、結の部分は 5.6% であった。各シーンの占める割合と入力となる物語テキストの総文数との積（小数点第 1 位で四捨五入して整数にする）が、各シーンにおける文の数となる。また 1 シーンの文の数が分かれば、分割点を 1 つに定めることができる。

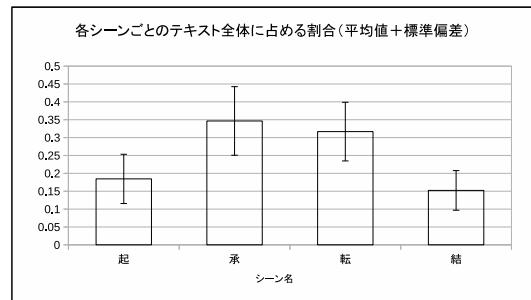


図 3: 各シーンの占める割合

また正解データを参照したところ、人間が起承転結の切れ目として認識するのは

- 文頭が「その時」・「やがて」・「つぎの（つぎの日、つぎの朝など）」で始まる箇所
- 文中に「ある日」というワードを含む箇所

が多いことが分かった。そこで、この条件に該当する箇所も分割点である可能性が高いとして検出した。しかし、この条件に該当する箇所が承・転・結のどの開始点であるかは分からないため、正解データから得られた分割点の情報を元にどこに該当するか判断する必要がある。これは正解データを参照し、各分割点との差の絶対値が最小になる部分に振り分けた（次節、手順 3 に該当）。

4.4 正解データの情報を援用した分割

作成した正解データから得られた分割点、及び特定ワードにより検出した分割点の情報を、吸引力による分割に取り入れる。分割の手順は以下の通りである。

手順 1

吸引力（4.1 節、4.2 節）による自動分割を行う（それぞれ手法 3-1、3-2 に対応）。

手順 2

手順 1 の結果を承・転・結の分割点のいずれかに振り分ける。まず手順 1 により得た分割点それぞれに対し、以下の値を算出しその絶対値を取る。

$$\text{差 } X = \text{手順 1 により得た個々の分割点} - Y \text{ の分割点}$$

ここで $\{X, Y\}$ は、 $\{1, \text{承}\}$ 、 $\{2, \text{転}\}$ 、 $\{3, \text{結}\}$ とする。差 X が最小の時の Y の分割点として振り分ける。ま

た差 X が正解データから判明した標準偏差より大きい場合、外れ値と見なして分割結果から除外する。

手順 3

全ての特定ワードについて以下の操作を行い、得られた分割点を手順 2 の結果に加える。下記の値を算出して絶対値を取り、差 X' が最小の時の Y の分割点として振り分ける。

差 $X' = \text{特定ワードにより検出した分割点} - Y$ の分割点

ここで $\{X', Y\}$ は、 $\{1', \text{承}\}$, $\{2', \text{転}\}$, $\{3', \text{結}\}$ とする。

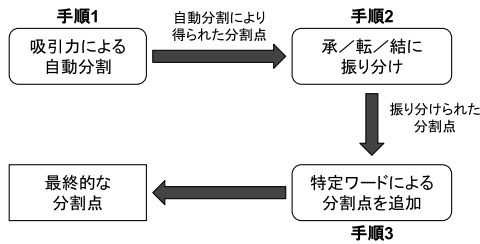


図 4: 手順 1~3

5. 実験結果と考察

5.1 実験結果

実装した各手法で、星新一ショート 20 篇に対し自動分割を行った結果を以下の 4 つの表に示す。評価の指標として F 値を用いた。

表 2: 手法 1: 吸引力による分割

平均の P 値	平均の R 値	平均の F 値	分割点の平均の個数
0.451	0.386	0.377	4.19

表 3: 手法 2: 対象範囲を狭めた吸引力による分割

平均の P 値	平均の R 値	平均の F 値	分割点の平均の個数
0.560	0.362	0.420	3.72

5.2 考察

結果的に、正解データの統計情報も利用した手法の方が F 値は高くなっている。その理由として、

- 正解データの分割点を参照することで、本来の分割点にそぐわない箇所を除くことができた
- 人間が分割点と感ずる箇所を加えることで、正解の検出数が増加した

という 2 点が考えられる。また吸引力については 2 種類の算出方法を試みたが、まず正解データの統計情報を用いない場合

表 4: 手法 3-1: 正解データの情報を援用した分割 (手法 1 の吸引力)

平均の P 値	平均の R 値	平均の F 値	分割点の平均の個数
0.799	0.541	0.620	3.72

表 5: 手法 3-2: 正解データの情報を援用した分割 (手法 2 の吸引力)

平均の P 値	平均の R 値	平均の F 値	分割点の平均の個数
0.829	0.513	0.608	3.25

は、算出方法を変更した手法の方が F 値はやや高くなっている。しかし正解データの統計情報を用いた場合は、算出方法を変更しない方が F 値が若干高くなっているため、算出方法変更後の方が有用であるとは断定できない。

他に全体的な特徴として、P 値が R 値よりも高いことが挙げられる。これは正解と見なす分割点の位置に幅を持たせたために、自動分割で検出されるシーンのうち正解の数が増加したためと推論される。恐らくこの幅を大きくすると F 値は高くなり、小さくすると低くなる。しかし幅を大きくしすぎると、人間の目を見た時に適切ではないと感じる分割点が含まれる可能性も高くなる。また検出された分割点の個数だが、理想とする出力数は 3 つ (承・転・結) である。それに対し、いずれの分割手法でも平均の分割点の個数は 3.25~4.19 と、極端に多すぎるようなことにはならなかった。

6. まとめ

本稿では、起承転結を持つ物語テキストを入力とした歌詞自動生成システムについての方針を述べた。また自動生成に必要な処理として構造抽出のアルゴリズムをいくつか考案し、実際の物語テキストに対し分割実験を行った。結果として、吸引力だけでなく人手によるシーン分割の統計情報も用いた方が F 値は高かった。また F 値の大きさ自体は最高値で 0.620 となり、この分割手法である程度は正しい分割点を検出できることが分かった。

今後の課題としては、物語テキストにおけるシーン自動分割手法の更なる改良、及び実際に歌詞を生成する処理の構築が挙げられる。

参考文献

- [1] 小林聡, 場・時・人に着目した物語のシーン分割手法, 情報処理学会研究報告, pp.25-30, 2007.
- [2] 赤石美奈, 文書群に対する物語構造の動的分解・再構成フレームワーク, 人工知能学会論文誌 21 巻 5 号, pp.428-432, 2006.
- [3] 森理美, ディスカッションマイニングにおける重要単語の抽出とその応用へ向けて, 平成 25 年度公立はこだて未来大学卒業論文, 2014.
- [4] 村井源, 松本育子, 佐藤知恵, 往住彰文, 物語構造の計量分析に向けて -星新一のショートショート物語構造の特徴-, 情報知識学会誌 Vol.21 No.1, pp.6-17, 2011.