

話題経路を用いた物語生成支援

Narrative Route for Story Generation

佐藤真 赤石美奈*1 堀浩一*2
 Makoto Sato*1 Mina Akaishi Koichi Hori

*1法政大学 Hosei University
 *2東京大学 University of Tokyo

In this paper, we propose a method for story generation. The method extracts a dynamics of topics in a story and applies the history of the dynamics to documents in order to create plots for new stories. The history of the dynamics is represented like a route on a map and the pieces of other stories are like landmarks. The method finds candidates for a new story plot on the route.

1. はじめに

物語のプロットを生成するために、物語の型を物語の概念空間における概念の変化とみなし、ある文書の型に基づき他の文書を再構成する手法を提案する。本手法では、文書を分割し、断片の特徴量に基づき地図上に配置し、断片を結んだパスを経路として、この経路を物語の型と見なす。そして他の物語の断片群を地図上に配置し、その経路に沿って断片を再構成することで、物語のプロットの生成支援を図る。

2. 話題経路の解析と利用

物語の断片を地図上にランドマークのように配置し、その経路を物語の型として、新しい物語を生成するためにその型を利用して既存の物語断片を再構成する手法を提案する。

たとえば小説における起承転結において文書におけるトピックの大小関係について次のような変化が想定できる：起から承かけて1つのトピックが大きくなる、転によって承で語られたトピックが一旦小さくすると同時に新しいトピックが発現して大きくなる、結によって承で扱ったトピックと転で扱ったトピックの両方が大きくなる。潜在意味分析の観点では、文書特徴量を次元圧縮して二次元の地図上に配置すると2つの軸が2つのトピックの大小が表していることになる。複数のトピックの大小の関係の変化を物語の起承転結の型に対応させれば、座標における経路を物語の型として見なすことができるだろう。さらに物語の型に沿って他の文書断片群をなぞれば、物語の異なる型における新しい物語を生成することができるだろう。

そこで、ある物語の型の抽出、抽出した型とあらかじめ蓄積しておいた断片群を用いた物語のプロットの生成を次の手順で実現する(図1)：(1)ある物語を分割して断片にする。(2)断片を空間上に配置してランドマークとする。ある物語における断片の出現順にランドマークをつないだ経路がある物語の経路とする。(4)あらかじめ蓄積しておいた物語断片群を空間上にランドマークとして配置する。(5)ある物語の経路に沿って配置された断片をつなぐ。再構成された断片列は、物語のプロットとなることが期待できる。

2.1 文書の分解と空間配置、経路

物語が記述された文書がある粒度に切り分け、分割した文書の特徴を量る。文書を適切な粒度に分割するためには、段落や連絡先：佐藤真，法政大学情報科学部，東京都小金井市梶野町3-7-2，042-387-4544，makoto.satoh.47@hosei.ac.jp

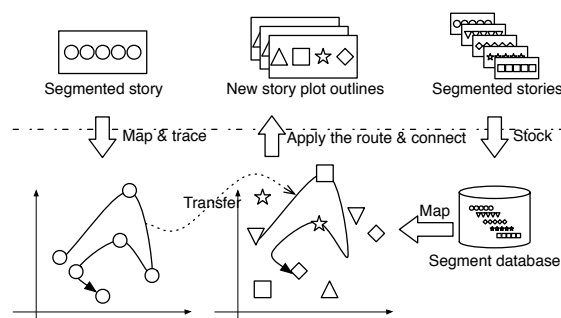


図1: 提案手法のプロセス

章のような明示的に与えられた区切りに応じて手作業で分割する方法、機械的にある一定の文字数や単語数で分割する方法、自動的に話題の切れ目を発見して分割する方法があるが、ここでは文書に形態素解析を施しストップワードを除去した後、機械的に一定語数ごとに区切る。

ある文書を n 分割した文書断片 d_i に出現した語 t_j の $tfidf(d_i, t_j)$ を要素とするベクトル \mathbf{d}_i を文書断片 d_i の特徴量と見なす。

文書断片ベクトルを次元削減を行い空間上に配置する。ここでは次元削減として多次元尺度構成法 (MDS) を用いる。MDS は文書断片の特徴量間の類似度の定義に基づき似ているものは近くに似ていないものは遠くに配置する手法である。

手順は次のとおりである。要素が距離の2乗と見なせる非類似度行列 S を求める。ここでは非類似度を \cos 距離で定義する。

$$S[i, j] = \text{dist}(\mathbf{d}_i, \mathbf{d}_j) = 1 - \frac{\langle \mathbf{d}_i, \mathbf{d}_j \rangle}{\|\mathbf{d}_i\| \|\mathbf{d}_j\|} \quad (1)$$

非類似度行列を中心化する。

$$P = -\frac{1}{2} \left(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}' \right) S \left(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}' \right) \quad (2)$$

中心化した非類似度行列 P の固有値 λ 、固有ベクトル \mathbf{v} を求め、固有値の大きい方から2,3個選び、対応する固有ベクトルを取り出す。このとき各固有ベクトルの要素値が座標となる。ただし2個の固有ベクトルの時は2次元、3個の固有ベクトル

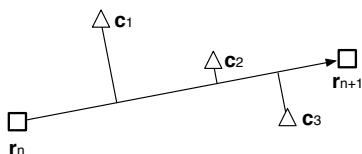


図 2: r_n から r_{n+1} への経路沿い候補の断片 c . ユークリッド距離に応じて候補順位を c_2, c_3, c_1 とする.

の時は 3 次元となる. たとえば断片 d_i の 2 次元座標 r_i は第 1 固有ベクトル v_1 , 第 2 固有ベクトル v_2 を用いて定義する.

$$r_i = [v_1[i], v_2[i]]^T \quad (3)$$

経路は座標列 $\{r_k\}_{k=1}^n = \{r_1, r_2, \dots, r_n\}$ で定義する.

2.2 話題経路の利用

解析した文書断片の経路を利用して, 他の文書の断片を連鎖させる. MDS を用いた次元削減においては文書の特徴量間の類似関係の相対的な変化が物語を特徴付けているものとみなせる. なお文書断片間の相対的な関係を保てればよいと考え, 経路座標二体に対して平行移動・回転・拡大縮小が許されるものとする.

他の文書の断片について前節同様の手順で座標を求めた後, 空間を重ね, ランドマークとする. 経路沿いにある断片を収集し連鎖させた断片列をプロットの候補とする. 図 2 では経路座標 r_n から r_{n+1} への経路沿いの候補の断片 c_1, c_2, c_3 を示す. 断片間を結ぶ線分とのユークリッド距離に応じて候補の順位を決める.

3. 実験

青空文庫の小説を対象に実験を行った.

まず, 文書の経路とランドマークを求め可視化する. 図 3 に芥川龍之介の羅生門と鼻について, MDS によって得られた 3 次元の座標に基づきプロットした例を示す. 羅生門に関しては渦を描くように物語が進んでいき, 鼻は話題が行ったり来たりしながら物語が進んでいくことがわかる.

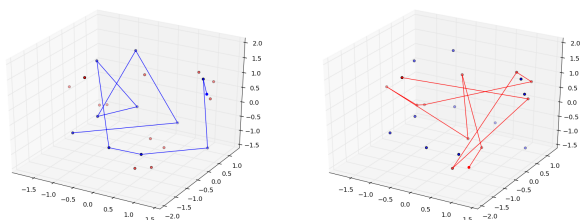


図 3: 左は羅生門の経路と鼻のランドマーク, 右は鼻の経路と羅生門のランドマーク.

次に, 羅生門, 鼻, ヴィヨンの妻, 人間失格, 或阿呆の一生の経路を求めた後, 走れメロスを再構成した. なお, 物語の起点の向きをなるべく一致させるため, 各座標を正規化した後, はじめの断片が同軸上に来るように回転した. 実験では, 400 字で区切った後, 形態素解析を施し, 名詞, 動詞, 形容詞を選択し特徴量に用いた. 走れメロスの断片は全 25 件で, 経路から近い 5 つ選択した. 表 1 に結果を示す. たとえば羅生門の

型で再構成されたプロットは 0,9,13,10,18 という順に並んだ. 各型によって異なる再構成結果が得られた.

表 1: 様々な物語の型を用いた走れメロスの再構成. プロットの数字はメロスの文書断片のインデックスを示す.

型	プロット
羅生門	0,9,13,10,18
鼻	0,13,14,22,4
ヴィヨンの妻	0,8,11,15,6
人間失格	7,8,22,13,17
或阿呆の一生	4,13,11,12,22

4. 関連研究

物語のようにデータを物語のように再構成する研究は数多くなされている. 物語の構造に関する研究に端を発した物語文法では物語のスキーマを設定することにより計算機によって物語を生成する [Thorndyke 77]. たとえば物語の登場人物のインタラクションのルールを決めておくことにより動的に物語を生成する [Theune 03]. また談話分析の分野においては, 記述された物語のレトリックに基づいて動的な意味の変化のモデルが提案されている [Lascarides 07]. 表層的な情報すなわち語の頻度や共起, 文脈のモデルにより情報を連鎖させる手法なども提案されている [Shahaf 10].

本研究では, 物語の表層的信息から潜在的な意味を解析するための手法を物語生成に適用して, 物語の表層的信息から話題を概念空間内のダイナミクスの履歴として抽出, 異なる概念空間で再利用することで新しい物語の構成の生成を支援する手法を提案した.

5. おわりに

物語経路を他の物語断片に適用して物語のプロットの生成の支援する手法を提案した. 本稿では多次元尺度構成法を用いて物語の文書断片を空間上に配置し経路を定義し, この経路を異なる同次元の異なる空間に当てはめ, 経路に沿って他の文書の断片を連鎖させた. 今後, 生成されたプロットに関する評価について考察し, また経路を観測とした文脈のダイナミクスのモデルを構築する予定である.

参考文献

- [Thorndyke 77] Thorndyke, Perry W. "Cognitive structures in comprehension and memory of narrative discourse." *Cognitive psychology* 9.1, pp. 77–110 (1977)
- [Theune 03] Theune, Marit, et al. "The virtual storyteller: Story creation by intelligent agents." In: TIDSE 2003, pp. 204–215 (2003)
- [Lascarides 07] Lascarides, Alex, and Nicholas Asher. "Segmented discourse representation theory: Dynamic semantics with discourse structure." *Computing meaning*. Springer Netherlands, pp. 87–124 (2007)
- [Shahaf 10] Shahaf, Dafna, and Carlos Guestrin. "Connecting the dots between news articles." *Proceedings of the 16th ACM SIGKDD* (2010)