

連想による物語性抽出と発想の飛躍による物語生成支援に関する研究

Story Generation Support by Extraction of the Story Nature due to Association and Idea Jump

木田 祐輝*¹
Yuki Kida

赤石 美奈*¹
Mina Akaishi

堀 浩一*²
Koichi Hori

*¹ 法政大学大学院情報科学研究科
Graduate School of Computer and
Information Sciences, Hosei University

*² 東京大学大学院工学系研究科
School of Engineering, University of Tokyo

The purpose of this research is story generation support that focuses on the relationship with the adjective and noun in documents. The proposed method maps to topics in a topic space to similar topics that appears in another topic space. Based on this mapping, terms in original story are replaced with other terms. It helps to discover new interpretation and support to make a new story.

1. はじめに

人が物語を創作する一つの方法として、書きたい事柄に関して調べ、さらに連想を広げ物語の発想を得る、ということが考えられる。この場合、得られた発想の中から物語の解釈を見つけ出し、そこから想像を膨らませることにより物語を作りだしていると考えられる。このことから、連想による物語性(物語の解釈)の抽出と、想像を膨らませ発想の飛躍を得ることによって物語生成が行われていると考えた。本研究ではこの考えを基に人の物語生成の支援を目指す。

まず、既存の物語作品の名詞とそれを修飾する語との関係性に着目した。例えば、ある文において「花」という名詞が「きれい」「美しい」といった形容詞によって修飾されている「きれいで美しい花」という表現について考える。次に「きれいで美しい空」という表現があるとする。この場合、「花」と「空」は同じ修飾語(形容詞)で修飾されているため、「きれい」で「美しい」という観点では近い関係があるといえる。しかしながら、概念的には「花」と「空」は近いとはいえない。そこで、形容詞との共起の類似度に基づき名詞とマッピングすることで発想の飛躍を提示し、これによって物語生成の支援を行おうと考えた。

本研究ではこのようにして、ある作家の作品群に登場する名詞を形容詞との共起性に基づき、別の作家の作品群に登場する類似した名詞へマッピングを行う。そうしてある作家の作品の名詞を別の作家の名詞に置き換えることで新しい解釈を発見する手助けを行い、これにより新しい物語を生み出す支援をする。これを用いた物語生成支援の検証を行う。

本稿では既存の物語として、星新一と芥川龍之介の作品群を使用した。星新一は1000篇以上のショートショートを書き、その多作さと質の高さから「ショートショート的神様」と呼ばれるほど著名である。また芥川龍之介は「羅生門」や「鼻」といった多くの傑作を残した小説家である。両者の作品群から類似する名詞候補を得ることとする。

2. 関連研究

語の出現依存度と吸引力による文書の語彙連鎖構造に着目し、文脈に基づき情報を分解・再構成する手法としてナラティブ連想情報アクセス[赤石 06]がある。また、語の吸引力を用い蓄積された電子文書から知識連繋のためのストーリー生成を行う手法としてトピックブリッジング[佐藤 11]がある。この手法はスタ

ートとゴールの2つのトピックを繋ぐためにはどのようなブリッジが適当であるかということ推定し、スタートからブリッジ、ゴールと連なるストーリー骨格を生成するものである。またこのトピックブリッジング手法を用いたアイデア発想支援システムとして[石川 11]がある。この研究では2つの異なるトピックをそれぞれ物語の始点・終点と捉えたトピックブリッジングを行い、宇宙開発というトピックと社会の利益というトピックを結び付ける予想外のトピックを発見する例が示されている。

本研究では、ある作家の作品群に登場するトピック(名詞)を別の作家の類似するトピックにマッピングし置き換えることで新しい解釈を発見する手助けを行う。これにより新たな物語を生み出す支援を目標とする。

3. 提案手法

3.1 提案手法の概要

はじめに、連想によって新しい解釈を発見する例を示す。ある人物A, Bが「向日葵」というトピックに対し「黄色い」「大きい」といった形容詞を連想したとする。この形容詞からAは「太陽」、Bは「麒麟」を連想したとする。Aからすれば、Bが「麒麟」を連想したことを意外に感じるだろう。このように人物間の連想の違いが新たな解釈を生み出すと考えられる。

図1に提案手法の概念図を示す。トピックを名詞としたとき提案手法は、あるトピック空間の名詞と形容詞との共起性を用い別のトピック空間の名詞へマッピングを行う。そしてあるトピック空間の名詞と、別のトピック空間の名詞を置き換えることで新しい解釈の発見を手助けし、新たな物語を生み出すことを支援する。提案手法では作家ごとの作品群文書を入力とし、文書に対して形態素解析による品詞分類を行い名詞と形容詞を抽出することで、ある作家のトピック空間の名詞を、別の作家のトピック空間の名詞へマッピングする。

3.2 トピックのマッピング

図2のような物語文書集合 S, S' に含まれる Topic(名詞)の集合を T, T' とおく。また S, S' に含まれる形容詞の集合を A とおく。 T, T', A のそれぞれの要素数を n, n', m とすると、それぞれの要素 t_i, t'_j, a_k は以下のように表される。

$$t_i \in T \quad (0 < i \leq n) \quad (1)$$

$$t'_j \in T' \quad (0 < j \leq n') \quad (2)$$

$$a_k \in A \quad (0 < k \leq m) \quad (3)$$

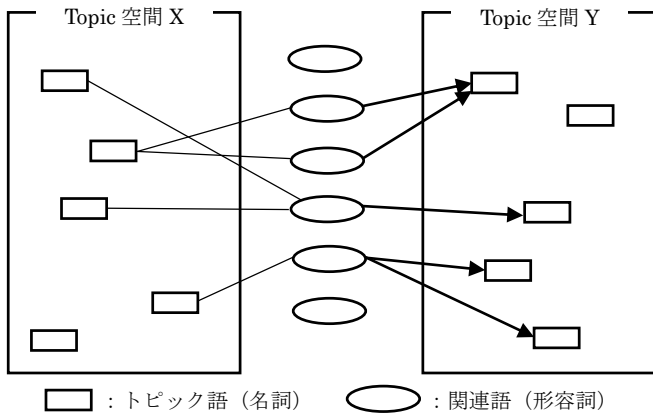


図1: 提案手法の概念図

図3のように、名詞 t_i と形容詞 a_k の共起頻度を $co(t_i, a_k)$ とする。また、名詞 t'_j と形容詞 a_k の共起頻度を $co(t'_j, a_k)$ とすると、単語ベクトル t_i, t'_j は次式(4)(5)で表される。

$$t_i = (co(t_i, a_1), co(t_i, a_2), \dots, co(t_i, a_k)) \quad (4)$$

$$t'_j = (co(t'_j, a_1), co(t'_j, a_2), \dots, co(t'_j, a_k)) \quad (5)$$

式(4)(5)を用いて、単語ベクトル同士の類似度をコサイン類似度で計算する。類似度計算を行う単語ベクトルを t_i, t'_j とすると類似度は、

$$\begin{aligned} \text{sim}(t_i, t'_j) &= \cos\theta \\ &= \frac{t_i \cdot t'_j}{|t_i||t'_j|} \end{aligned} \quad (6)$$

で表される。この類似度を用いて S に含まれるトピックを S' のトピックにマッピングする。この場合のトピックをマッピングしたリスト(TopicMappingList)は図4のようなになる。

3.3 物語文書の要約文生成

図5は物語文書の要約文を生成しトピックを置き換え、そこから人が新しい解釈を発見することで新しい物語を生み出す手順を表したものである。

物語文書の名詞をマッピングした名詞に置き換える際、物語文書の全文に対して変換を行うと長い文章間の文脈を考慮することができず、また日本語として意味の通らない文章となってしまうことがある。そこで物語文書の各文の重要度を計算し、重要度の高い文を抽出することで要約文を生成する。この要約文の名詞をマッピングした名詞に置き換え出力とすることで変換した文章の内容理解の補助を目指す。

物語文書の各文の重要度を計算し、重要度の高い文を抽出することで要約文を生成する。これには文章の特徴を表すキーワードを発見して重要文を抽出する展望台システム[砂山 01]という手法を用いる。この手法は、単語の出現頻度による周辺キーワード、周辺キーワードと同時に出現する割合の中心キーワード、中心キーワードと同時に出現する割合の高い特徴キーワード、この3つのキーワードの評価値を用い各文の重要度を計算し要約文を生成する。

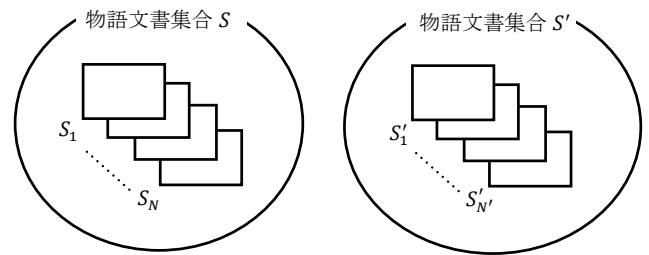


図2 物語文書集合の概念図

	a_1	a_2	...	a_k	...	a_m		a_1	a_2	...	a_k	...	a_m
t_1							t'_1						
t_2							t'_2						
\vdots							\vdots						
t_i				$co(t_i, a_k)$			t'_i				$co(t'_i, a_k)$		
\vdots							\vdots						
t_n							t'_n						

図3 名詞と形容詞の共起頻度リスト

	類似度：大 ↔ 小				
t_1	t'_{11}	t'_{12}	t'_{1x}
t_2	t'_{21}	t'_{22}	t'_{2x}
\vdots					
t_n	t'_{n1}	t'_{n2}	t'_{nx}

図4 TopicMappingList

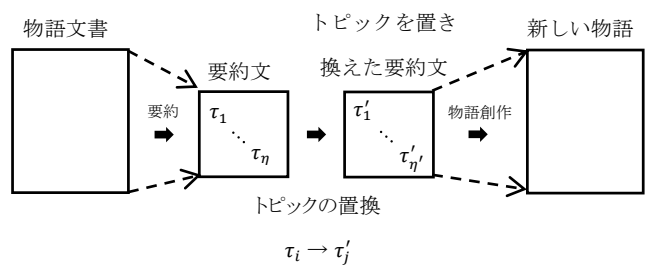


図5 要約文生成とトピックの置換

4. 実験と考察

提案手法による物語生成支援の検証実験を行った。実験に使用する物語文書集合は星新一の短編集「ボッコちゃん」に含まれる50作品と、芥川龍之介の長編、短編小説を含めた100作品を使用した。今回の実験では、TopicMappingListの単語で置き換える物語文書として星新一の短編集「ボッコちゃん」内のショートショート「悪魔」を使用した。

文書内で出現頻度が高い語は、その文書内での重要度が高いことを意味している。そこで、物語文書内の上位5つの名詞を TopicMappingList の名詞で置き換えることとし、物語文書内の名詞 τ_i を、TopicMappingList の名詞 τ'_j で置き換える際に、以下の2通りの方法で置き換え、結果を比較した。

- ① 自動置換: τ_i と最も類似度の高い名詞を τ'_j とする
- ② 手動置換: τ_i と $\tau_1 \sim \tau_n$ の関係を考慮しつつ、 τ_i と類似度の高い名詞の中から人が適切だと判断した名詞を τ'_j とする(置き換えるかどうかは任意)

また単語を置き換えた文書を比較する際、以下の3通りの文書で比較を行った。

- (i) 原文書: 要約を行わない物語文書
- (ii) 要約₁₀: 物語文書から重要度の高い10文を抽出した要約文
- (iii) 要約₄: 物語文書から重要度の高い4文を抽出した要約文

4.1 ショートショート「悪魔」での実験結果と考察

星新一のショートショート「悪魔」で提案手法による単語の置き換えをした結果を以下に示す。

はじめに、ショートショート「悪魔」のあらすじを説明する。エス氏が凍った湖の上で釣りをしていると、古いツボを釣り上げた。そのツボの中から悪魔が出てくる。悪魔がエス氏の願い事を叶えるというのでエス氏は金銭を要求する。悪魔は金貨を出す。欲にかられたエス氏は「もつともつ」とさらなる金貨を要求する。次第に湖の氷の上に大量の金貨が積まれていき、やがてその重さで氷が割れ悪魔と大量の金貨は湖の底に沈んでいってしまう、というものである。

表1は「悪魔」の文書内で出現頻度が高い上位5つの名詞とその出現頻度、そして星新一と芥川龍之介の物語文書集合内でその名詞と共起関係にある形容詞のリストである。

表2は、星新一の文書「悪魔」の出現頻度が高い上位5つの名詞と、芥川龍之介の物語文書集合の名詞とのマッピング結果 (TopicMappingList) である。

表3は要約₄の要約文を表にまとめたものである。置換対象となる語(名詞)をアンダーラインで示し、さらに出現頻度の高い上位5つの語を太字で示す。

はじめに、自動置換の方法で原文書、要約₁₀、要約₄の文書の名詞を置き換えた。名詞「金貨」の TopicMappingList の最も類似度の高い名詞は「鶴」であり、この名詞で置き換えた場合、「一枚の金貨を差し出した」などの文において意味が通らない文となってしまう。これは単語の概念や係り受け関係などを考慮することなく置き換えているためである。同様に「悪魔」と「話し声」、「氷」と「色目」、「ツボ」と「季節」、「目」と「倶楽部」の置き換えにおいても意味の通らない文となることが多い。

自動置換で要約₄の名詞を置き換えた結果を表4に示す。この場合4文全てを考慮し意味を取ることは難しいが、自動置換 II と自動置換 III の文はそのままで読むことができなくもない。エス氏が鶴を大量に要求することでどうなるのか、という解釈のもとに想像を膨らませて行けば新しい物語を生み出すきっかけとなるかもしれない。

以上のことから、単語の概念や係り受け関係などを考慮しない自動置換の方法では名詞を置き換えることによって意味の通

表1 「悪魔」内の頻出名詞と共起関係にある形容詞リスト

名詞	出現頻度	物語文書集合内で共起関係にある形容詞
金貨	8	長い, 高い, うれしい, まぶしい...
悪魔	8	高い, 楽しい, うれしい, ひどい...
氷	5	高い, 厚い, 丸い, かん高い
ツボ	4	高い, 古い, かん高い
目	4	いい, 早い, すばらしい, 長い...

表2 「悪魔」の TopicMappingList

t_i	t'_j	類似度: 大 ↔ 小
金貨	鶴, 機関, 診察, 髯, 烏帽子, 靴下...	
悪魔	話し声, 赤面, 勇者, 家内, 占い, 到着...	
氷	色目, 確信, 踵, 土手, 紙切れ, 匂い...	
ツボ	季節, 季節外れ, むぎ, 麦藁, 和服...	
目	倶楽部, 立腹, 超人, 身代り, 急, 屋敷...	

表3 要約₄の要約文

ID	要約 ₄
I	<u>悪魔</u> は <u>氷</u> の穴に、ちよつと <u>手</u> をつつこんだかと思うと、一枚の <u>金貨</u> をさし出した
II	<u>首</u> をかしげながら、エス氏が <u>手</u> にとってみると、 <u>本物の金貨</u> にまちがいない
III	エス氏は何回もねだり、 <u>悪魔</u> はそのたびに <u>金貨</u> を出してくれた
IV	やつとたどりつき、ほつとしてふりかえってみると、 <u>氷</u> は大きな音をたてて割れ、 <u>金貨</u> も <u>ツボ</u> も、かん高い <u>笑い声</u> をあげている <u>悪魔</u> も、みな <u>湖の底</u> へと消えていった

表4 自動置換で名詞を置き換えた要約₄

ID	自動置換: 要約 ₄
自動置換 I	<u>話し声</u> は <u>色目</u> の穴に、ちよつと <u>手</u> をつつこんだかと思うと、一枚の <u>鶴</u> をさし出した
自動置換 II	<u>首</u> をかしげながら、エス氏が <u>手</u> にとってみると、 <u>本物の鶴</u> にまちがいない
自動置換 III	エス氏は何回もねだり、 <u>話し声</u> はそのたびに <u>鶴</u> を出してくれた
自動置換 IV	やつとたどりつき、ほつとしてふりかえってみると、 <u>色目</u> は大きな音をたてて割れ、 <u>鶴</u> も <u>季節</u> も、かん高い <u>笑い声</u> をあげている <u>話し声</u> も、みな <u>湖の底</u> へと消えていった

る文章にすることは難しいといえる。しかし、短い要約文を生成することで意味の通る文を見つけやすくなる可能性はある。その文から新しい解釈を発見し想像を膨らませることで新しい物語を生み出すきっかけとなるならば、自動置換の方法で物語生成の支援を行える可能性は残されているといえるだろう。

また、表4のような要約文を物語のプロットと見ることができれば、自動置換の方法で物語プロットの自動生成を行える可能性も考えられる。

次に、手動置換の方法で原文書、要約₁₀、要約₄の文書の名詞を置き換えた。名詞 τ_i と $\tau_1 \sim \tau_n$ の関係を考慮すると、「氷」と「湖」は同じ場所であるという関係があり、「ツボ」と「つり針」には「ひっかかる」という動詞が、「悪魔」と「金貨」は「取り出す」という動詞が係っていることなどが分かる。

原文書の場合は、文章全体の文脈を把握しつつ、出現頻度の高い語の関係性も考慮に入れなければならないため、TopicMappingList の名詞を選択することは難しい。また意味の通る文となるような名詞を選択したとしても、名詞の置き換えが起きない文も数多く存在するため元の物語文書のイメージが強くなり、新しい解釈を得ることは難しいといえる。

要約₁₀の文書の場合は、原文書ほど全体の文脈を把握する必要がないため TopicMappingList の名詞を選択することは難しくなく、ここで、「悪魔」という名詞に対して TopicMappingList の「勇者」という名詞がマッピングされていることに注目する。「悪魔」という名詞には善悪でいうところの「悪」のイメージが付きまとう。それに対し「勇者」という名詞には「正義」や「善」のイメージがあり、これは全く反対のイメージを持った名詞がマッピングされていることを示している。この物語文書の重要な登場人物である「悪魔」を正反対のイメージを持った「勇者」に置き換えた場合どうなるのか、という新しい解釈を見出すことができるだろう。

要約₄の文書の場合は、原文書、要約₁₀の文書の場合と比較すると文章量が少ないため TopicMappingList の名詞を選択することはそれほど難しくなく思われる。

要約₄の文書の名詞を置き換えた結果を表5に示す。この文書では「金貨」を「靴下」、「悪魔」を「家内」、「氷」を「土手」、「ツボ」を「和服」に置き換えた。名詞「目」は要約文内には含まれていないため置き換えを行っていない。

要約文を生成するために使用している展望台システムという手法では名詞や動詞の出現頻度だけでなく、頻出語と同時に出現する割合の高い単語にも高い評価値を与えているため、重要文として抽出した文の中に「目」のような出現頻度の高い名詞が含まれていない場合も存在する。このことから、単純に出現頻度の高い上位の名詞に焦点を当てるのではなく、要約生成において重要度が高いと評価された単語の置き換えも考慮に入れる必要があるといえるだろう。

表5は全体として意味の通る文章となっていることが分かる。また、なぜ家内は土手の穴から靴下を取り出したのか、なぜ土手の穴に靴下があるのか、なぜエス氏は「靴下」を何回もねだるのか、そういった新しい解釈を発見することができるといえるだろう。これを基に想像を膨らませることで新しい物語の生成を支援できるのではないだろうか。つまり、本手法によりメインストーリーを生成した後、伏線となるサブストーリーを書き足すことで物語を完成させていくことにつながれると考えられる。

表5の文章を物語のプロットと見た場合、手動置換IVのオチに当たる部分を工夫することで新しい物語を生み出せるのではないだろうか。このことから、手動置換の手法を用いることで物語プロットの生成支援を行えることも考えられる。

以上のことから、自動置換、手動置換のどちらの手法においても新しい解釈を発見することができていることがわかる。また物語文書の要約文を用いることで文章の理解や名詞の置き換えが容易となり、新しい解釈を発見しやすくなることが伺える。名詞を置き換えた要約文を物語のプロットと見ることができればプロットの自動生成や生成支援を行える可能性も考慮できる。

表5 手動置換で名詞を置き換えた要約₄

ID	手動置換:要約 ₄
手動置換 I	家内は土手の穴に、ちょっと手をつつこんだかと思うと、一枚の靴下をさし出した
手動置換 II	首をかしげながら、エス氏が手にとってみると、本物の靴下にまちがいない
手動置換 III	エス氏は何回もねだり、家内はそのたびに靴下を出してくれた
手動置換 IV	やっとなどつき、ほっとしてふりかえってみると、土手は大きな音をたてて割れ、靴下も和服も、かん高い笑い声をあげている家内も、みな湖の底へと消えていった

5. まとめ

ある作家の作品群に登場するトピックを別の作家の類似するトピックにマッピングし置き換えることで新しい解釈を発見する手助けを行い、物語生成を支援する手法を提案した。トピックを名詞としたとき形容詞との共起性に基づき類似する名詞へマッピングを行う。

提案手法の検証として星新一の物語文書内の名詞と芥川龍之介の作品群の名詞を置き換える実験を行った。名詞を置き換える方法として①最も類似度の高い名詞と置き換える方法(自動置換)、②文脈を考慮し人が適切だと判断して名詞を置き換える方法(手動置換)の2通りを、物語文書と要約文に対して行った。その結果、両方の方法で新しい解釈を発見することができた。また要約文を用いて名詞を置き換えることで新しい解釈を発見しやすくなる可能性があることがわかった。さらに名詞を置き換えた要約文を物語プロットとみなすことができれば、自動置換の方法ではプロットの自動生成が行える可能性があり、手動置換の方法ではオチに当たる部分を工夫することでプロットの生成支援が行えるのではないかとこの考察に至った。

今後は名詞と動詞との共起頻度や係り受け解析なども考慮に入れることでトピックのマッピングを改良し、その上で物語生成支援の検証を行っていく。

謝辞

本研究の遂行にあたり、ご協力頂いた星ライブラリ/新潮社様に心より感謝いたします。

参考文献

- [赤石 06] 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, Vol. 21, No. 5, pp. 428-438, 2006.
- [佐藤 11] 佐藤 真, 赤石 美奈, 堀 浩一: トピックブリッジングによるストーリー生成支援, 2011 年度人工知能学会全国大会, 2011.
- [石川 11] 石川 雄基, 佐藤 真, 堀 浩一, 赤石 美奈: 宇宙開発と社会との新たな連関を探るトピックブリッジング手法, 2011 年度人工知能学会全国大会, 2011.
- [砂山 01] 砂山 渡, 谷内田 正彦: 文章の特徴を表すキーワードを発見して重要文を抽出する展望台システム, 電子情報通信学会論文誌, Vol. J83-D1, No. 2, pp. 146-154, 2001.