

言語モデルによる物語中のイベントの顕現性推定

大竹 孝樹¹ 横井 祥^{1,2} 井之上 直也^{1,2} 高橋 諒^{1,2} 栗林 樹生^{1,3} 乾 健太郎^{1,2}

¹ 東北大学 ² 理化学研究所 ³ Langsmith 株式会社

{takaki, yokoi, naoya-i, ryo.t, kuribayashi, inui}@ecei.tohoku.ac.jp

1 はじめに

我々は物語の類似性を自然に認識する能力を持っている。例えば「恵まれない境遇の女性が幸運を掴み成功を手に入れる」という物語は、我々に童話『シンデレラ』との類似を思い起こさせる。実際に『シンデレラ』と類似する物語は地域や時代を超えて存在することが知られており [1, 23], こういった物語の類似性, 類型を対象にした研究は民俗学, 物語論などの人文学 [1, 16, 23] 及び自然言語処理分野 [5, 8–10, 13, 14, 19, 20] で盛んに行われてきた。

民俗学や物語論の分野では, イベントの**顕現性** (顕著さ, 重要さ) に着目して物語の類型分類・分析を行う研究が多く存在する [1, 16, 23]. 物語には様々なイベント (できごと) が登場するが, それらが全て同じ顕現性をもつわけではない。例えば, 童話『シンデレラ』において「シンデレラが王子に見初められる」というイベントは物語において重要な役割を果たし, 物語を特徴づける顕現性の高いイベントだが, 「シンデレラが井戸で水を汲む」というイベントはそうではない。民俗学および物語論における物語の類型分類・分析の研究により, 物語の類似性の認識における顕現性の高いイベントの重要性が示唆されており, 物語の類似性を計算機によって計算する際にもこういったイベントの顕現性を考慮することが重要であると考えられる。しかしながら, 自然言語処理の分野では物語の類似性を計算する際のイベントの顕現性の計算方法についてはほとんど検証・議論がなされてこなかった。そこで本研究では物語の類似性の計算に向け, イベントの顕現性の計算方法に着目する。本研究の貢献は以下の通りである。

- 物語論における物語の類型分類・分析の研究に基づき, 言語モデルを利用したイベントの顕現性の計算方法を提案した。
- 提案法の有効性を実験により検証した。

2 関連研究

人文学の知見を足がかりとしてイベントの顕現性の計算方法を検討するため, 本節では物語論及び民俗学においてイベントの顕現性という概念がどのように取り扱われてきたのかを概観する。

2.1 物語論におけるイベントの顕現性の概念: 機能

物語論では, 物語の構造分析を目的とした様々な研究が行われてきているが, それらに大きな影響を与えた初期の研究にロシアの民俗学者, Vladimir Propp による研究 [16] がある。Propp はロシアの魔法昔話を分析する中で**機能 (function)** と呼ばれる概念を見出し, 機能の列に注目することでロシアの魔法物語に共通の構造を発見した [16]. 機能は

“登場人物の行為で, しかも, 筋=出来事全体の展開過程にとって当の行為がもちうる意義 (位置) という観点から規定された登場人物の行為” [17]

と定義され, Propp はロシアの魔法昔話には 31 個の機能が存在すると分析した。

- 機能 1 家族の成員のひとりが家を留守にする (留守)
- 機能 2 主人公に禁を課す (禁止)
- 機能 3 禁が破られる (違反)
- ...
- 機能 31 主人公は結婚し, 即位する (結婚)

それぞれの物語がこれら全ての機能を含むわけではなく, 複数の機能が欠落することもあるが, Propp はこれらの機能は分析対象とした全てのロシアの魔法昔話に同じ順序で現れることを発見した。また, フランスの文学理論家・哲学者である Roland Barthes は機能の中でも物語の進行における重要度が高いものを**枢軸機能体 (cardinal function)** と呼び, Propp による物語の構造分析の理論をさらに深めた。枢軸機能体は

“物語の行動にとって論理的に本質的なものであって、そのひとつでも削除されれば、その因果的・年代記的な首尾一貫性は損なわれてしまう” [15]

ものと定義される。Barthes は [3] の中で小説テキストを例に、この概念に対するより詳細な説明を与えている。本稿では Propp 及び Barthes の研究を足がかりとして物語におけるイベントの顕現性の計算方法を検討する。

2.2 民俗学における民話分類の鍵: モチーフ

民俗学では、様々な地域に伝わる民話を収集し、それらの分類体系を構築する研究が行われてきた [1, 23]. Aarne-Thomson-Uther type index (以下 ATU 分類) [1, 23] は最も広く用いられる民話の分類体系の一つであり、様々な地域から収集された民話に対し、最も細かい粒度では約 2,400 のクラスに分かれる分類が行われている。分類は階層構造をなしており、上位の階層では「魔法物語」や「動物物語」などトピックによる粗い分類がなされ、最下層では「The Sleeping Beauty (眠れる森の美女)」や「Outcast Animals Find a New Home (ブレイメンの音楽隊)」のような詳細な分類がなされる。

ATU 分類を含む多くの民話の分類体系は、**モチーフ (motif)** に基づいた分類を行っている。モチーフの明確な定義は研究者によって揺れ動いているが、概ね“伝承の中に生き残る力を持った説話の中の最小の要素”とされる [22, 25]. シンプルな物語は 1 つのモチーフからなることもあるが、多くの物語は複数のモチーフから構成される [2].モチーフの例としては、たとえば「Tabu: eating food of certain person (誰かの食べ物を食べてしまうという禁忌を犯す)」や「Crimes punished (罪が罰せられる)」が挙げられる。モチーフの多く*1は単一の出来事(イベント)であり [25], これらは物語を特徴づける顕現的なイベントである。ATU 分類は Propp の研究と並び、イベントの顕現性に注目して物語の類型・類似性を分析した人文学における研究の例である。

2.3 自然言語処理分野における既存研究

自然言語処理の分野では、物語の類似性を計算する際に上述したイベントの顕現性を明示的に考慮する研究はほとんど行われてこなかった。多くの研究は語彙レベルの類似性に基づくもの [9, 14, 20] であり、構造的な類似性 [10, 19] や登場人物 [5] に焦点を当てたものも一部存在する。[13] は (S,V,O) の 3 つ組で表現されるイベントの TF-IDF 値を素性の一つとすることで実質的にイベントの顕現性を考慮した物語の類似性尺度を設計

*1単一のできごとの他に行為者や行為の背景を含む [25]

しており、物語の類似性計算にわずかに有効であることを報告している。

3 イベントの顕現性の計算方法

3.1 顕現性の高いイベントが満たすべき要件

イベントの顕現性の計算方法を検討するため、物語論における Propp の機能及び Barthes の枢軸機能体の定義を足がかりに、顕現性の高いイベントが満たす要件を定める。Propp が述べる「物語の筋の展開において意義を持つイベント」が仮に物語テキスト上で削除された場合を考えると、そのイベントが物語の筋の展開において意義を持つが故に、物語全体としての首尾一貫性 (coherence) は大きく損なわれてしまうと考えられる。例えば、『シンデレラ』のあらすじの一部*2は

- シンデレラが舞踏会へ行く
- シンデレラが姉達に会場で挨拶をする
- 会場でシンデレラが王子に見初められる
- ...
- シンデレラが王子と結婚する

というものが、ここで「会場でシンデレラが王子に見初められる」という物語の筋の展開において大きな意義を持つイベントが物語テキスト上から削除されると、物語の筋書きは以下ようになる。

- シンデレラが舞踏会へ行く
- シンデレラが姉達に会場で挨拶をする
- 会場でシンデレラが王子に見初められる <削除>
- ...
- シンデレラが王子と結婚する

削除後の物語は「会場でシンデレラが王子に見初められる」→「シンデレラが王子と結婚する」や「シンデレラが舞踏会へ行く」→「会場でシンデレラが王子に見初められる」など、複数の意味レベルの関連性が失われ、物語全体としての首尾一貫性が大きく損なわれたものになってしまう。一方で、「シンデレラが姉達に会場で挨拶をする」という物語の筋の展開において大きな意義を持たないイベントが削除されると、物語の筋書きは以下ようになる。

- シンデレラが舞踏会へ行く
- シンデレラが姉達に会場で挨拶をする <削除>
- 会場でシンデレラが王子に見初められる
- ...
- シンデレラが王子と結婚する

削除後の物語は、物語全体としての首尾一貫性が大きく

*2説明のため、あえて一般的に『シンデレラ』のあらすじには登場しないが、物語に存在するイベントを含めている

損なわれることはない。

以上を踏まえ、本研究では顕現性の高いイベントの要件を「物語テキスト上で削除されると物語全体としての首尾一貫性が大きく損なわれるイベント」と定める。これは、Barthesの枢軸機能体の定義の言い換えともみなせる。

3.2 イベントの顕現性の計算方法

近年では、膨大なラベルなしテキストで事前学習された言語モデルが、物語生成において最高性能を達成しているモデル [6] と比較して、物語テキストの首尾一貫性をより捉えていると報告されている [21]。そこで本研究では、イベントの削除による物語の首尾一貫性の変化を言語モデルによって捉える。具体的には、言語モデルが物語テキストに与える尤度をその物語の首尾一貫性を反映した値と見なし、オリジナルの物語と、あるイベントが削除された物語に対する尤度の比によってイベントの顕現性を計算する。

物語からイベントを“削除”する方法 物語においてあるイベントを削除する（該当するイベントの情報を落とす）方法にはいくつか選択肢が存在する。イベントを表現する動詞を削除する方法、主語や目的語を考慮し述語項構造の単位で削除する方法、イベントを含む文を削除する方法などが考えられ、“削除”の方法についても、テキスト上から文字列を削除するのではなく、イベントを表現する動詞を“do”などの一般的なものに置換する方法も考えられる。本研究ではイベントを含む文を削除することで該当するイベントを物語上から削除する。この方法はオリジナルの物語に対して正確に対象のイベントの情報のみが削除された物語の尤度比を計算できているわけではなく、あくまで近似的な計算方法である。

提案法 以下、記号 $S_{\{i:j\}}$ で文の列 $(S_i, S_{i+1}, \dots, S_j)$ を表す。また、文 S_i に含まれる単語を $(w_1^{(i)}, \dots, w_{|S_i|}^{(i)})$ と表す。

オリジナルの物語 $S_{\{1:n\}}$ と、顕現性を計算したい文 S_k を取り除いた物語 $S_{\{1:n\} \setminus \{k\}}$ について、 S_k から先の物語 $S_{\{k+1:n\}}$ の生成尤度をそれぞれ求め、その比を文 S_k の顕現性スコアとする：

$$\frac{1}{|S_{\{k+1:n-\ell'\}}|} \log \frac{P(S_{\{k+1:n-\ell'\}} | S_{\{1+\ell:k\}})}{P(S_{\{k+1:n-\ell'\}} | S_{\{1+\ell:k\} \setminus \{k\}})} \quad (1)$$

$$= \frac{1}{|S_{\{k+1:n-\ell'\}}|} \left[\sum_{i=k+1}^{n-\ell'} \log P(S_i | S_{\{1+\ell:i-1\}}) - \sum_{i=k+1}^{n-\ell'} \log P(S_i | S_{\{1+\ell:i-1\} \setminus \{k\}}) \right]. \quad (2)$$

表1: データセットの基本統計量

| 物語の総数 | 15 |
|--------------------|--------|
| 1つの物語に含まれる文の数 (平均) | 86.8 |
| // 単語の数 (平均) | 1257.5 |
| // 機能の数 (平均) | 11.3 |

ただし $P(S_i)$ は単語の生成確率の積：

$$\log P(S_i | \text{context}) = \log P(w_1^{(i)} | \text{context}) + \sum_{j=2}^{|S_i|} \log P(w_j^{(i)} | \text{context}, w_1^{(i)}, \dots, w_{j-1}^{(i)}). \quad (3)$$

$|S_{\{i:j\}}|$ は (S_i, \dots, S_j) 全体に含まれる単語数の総和を表す。 ℓ' および ℓ は言語モデルに与えることのできる単語数の上限 L によって決める閾値で、 $|S_{\{k+1:n-\ell'\}}|$ 、 $|S_{\{1+\ell:k\}}|$ がそれぞれ $L/2$ 以下かつ最大の値となるよう定める。

各物語における最後の文に対する顕現性スコアも他の文と同様に計算できるよう、あらかじめ各物語の末尾にテキストの区切りを表す特殊トークンを追加する。最後の文の顕現性スコアは、この特殊トークンの生成尤度（直後に物語が終了する尤もらしさ）を介して計算することができる。

4 実験

顕現的なイベントがアノテーションされた物語コーパスを用い、提案手法の有効性を検証する。

4.1 実験設定

データセット 15編のロシアの魔法昔話に対して Propp の「機能」がアノテートされたデータセットである ProppLearner [7] を用いる。コーパス中では動詞（句）に対して機能のアノテーションがなされているため、前処理として、当該の動詞（句）を一部でも含む文を顕現性の高い文とみなす。提案手法を含む各システムは、文単位で顕現性が高いか否かを推定すればよい。データセットの基本統計量を表1に示す。各物語に対して平均 86.8 個の文が含まれており、データセット全体で約 1300 の文を含む。

提案法 顕現性スコアの予測に用いる言語モデルとして、事前学習済みの GPT2 [18] を使用する*3。

ベースライン手法 比較対象として、各文に対する顕現性スコアをランダムで予測するベースラインを用いる。

評価尺度 今回の実験設定を物語を構成する各文を顕現性の高い順にランキングする問題と見なし、同様の設定である情報検索の2値適合性に基づくランク付き検

*3実装は transformers [24] の GPT2LMHeadModel を使用する

表2: 実験結果

| | MAP |
|------------|-------|
| ランダムベースライン | 0.212 |
| 提案法 | 0.255 |

索の標準的な評価尺度である Mean Average Precision (MAP) [12] を用いる。

4.2 実験結果

ランダムベースラインとの比較 MAP による評価の結果を表2に示す。ランダムベースラインに比べ提案手法が約 0.043 ポイント上回った。提案法は顕現性の計算手法として一定程度のポテンシャルを持つものと考えられる。

物語の末尾の文に対する顕現性スコア 各物語における最後の文の顕現性スコアが、極端に高い、もしくは極端に低い値をとる傾向が見られた。提案法では、各文に対して、続く物語の尤度比のトークン平均を当該の文の顕現性のスコアとする。最後の文に関しては「続く物語」が物語末尾に付与した特殊トークンのみであり、平均される値の個数が極端に小さいため、分散が大きくなったものと考えられる。対策として、順方向の言語モデルと逆方向の言語モデルを併用することで、当該の文の物語内の相対位置に頑健な手法に変更することが考えられる。

文の長さとの顕現性スコアの相関 提案法は、長い文に対して高い顕現性スコアを示す傾向があった（文が含むトークン数と顕現性スコアのピアマンの順位相関係数は 0.209 であった）。長い文は機能となるイベントを含む可能性が高く、その結果、当該のその文を削除した際に続く物語の生成尤度が下がりやすかったと考えられる。

5 おわりに

本稿では物語論における Propp や Barthes の研究に基づき、言語モデルを利用してイベントの顕現性を計算する方法を提案した。顕現性の高いイベントがアノテーションされた物語コーパスを利用した実験により、提案法はイベントの顕現性の推定にある程度有効である可能性が示唆された。

今後の課題として、順方向言語モデルと逆方向言語モデルの併用や、イベントを削除する方法の工夫により提案法を改善し、TF-IDF や要約などで用いられるイベントの顕現性の計算手法との比較を行いたい。また、テキストの首尾一貫性の計算モデルの既存研究 [4, 11] の知見を取り入れ、首尾一貫性の計算方法の改善に取り組みたい。さらに、イベントの顕現性スコアを考慮した物語の類似性尺度を設計し、イベントの顕現性を考慮するこ

とが物語の類似性計算においてどの程度有効なのか検証したい。

謝辞. 本研究は、JST, CREST, JPMJCR1301 の支援を受けたものである。

参考文献

- [1] A Aarne and S Thompson. *The types of the folktale: a classification and bibliography*. FF communications. Suomalainen Tiedeakatemia, Academia Scientiarum Fennica, Helsinki, 1961.
- [2] Dee L Ashliman. *Folk and Fairy Tales: A Handbook*. ABC-CLIO, 2004.
- [3] Roland Barthes and Lionel Dussit. “An Introduction to the Structural Analysis of Narrative”. In: *New Literary History* 6.2 (1975), p. 237.
- [4] Regina Barzilay and Mirella Lapata. “Modeling Local Coherence: An Entity-Based Approach”. In: *Computational Linguistics* 34.1 (2008), pp. 1–34.
- [5] Snigdha Chaturvedi et al. “Where Have I Heard This Story Before? Identifying Narrative Similarity in Movie Remakes”. In: *NAACL Association for Computational Linguistics*, 2018, pp. 673–678.
- [6] Angela Fan et al. “Hierarchical Neural Story Generation”. In: *ACL Association for Computational Linguistics*, 2018, pp. 889–898.
- [7] Mark A Finlayson. “ProppLearner: Deeply annotating a corpus of Russian folktales to enable the machine learning of a Russian formalist theory”. In: *Digital Scholarship in the Humanities* 32.2 (2015), pp. 284–300.
- [8] Mark Alan Finlayson. “Learning narrative structure from annotated folktales”. PhD thesis. Massachusetts Institute of Technology, Cambridge, MA, {USA}, 2012.
- [9] Lisa Friedland and James Allan. “Joke retrieval: recognizing the same joke told differently”. In: *CIKM*. 2008, pp. 883–892.
- [10] Victoria Anugrah Lestari and Ruli Manurung. “Measuring the Structural and Conceptual Similarity of Folktales using Plot Graphs”. In: *LaTeCH Association for Computational Linguistics*, 2015, pp. 25–33.
- [11] Jiwei Li and Dan Jurafsky. “Neural Net Models of Open-domain Discourse Coherence”. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2017, pp. 198–209.
- [12] Christopher D Manning et al. “Evaluation in information retrieval”. In: *Introduction to Information Retrieval*. Cambridge University Press, 2008, pp. 139–161.
- [13] Dong Nguyen et al. “Folktale Classification Using Learning to Rank”. In: *ECIR*. 2013, pp. 195–206.
- [14] Dong Nguyen et al. “Using Crowdsourcing to Investigate Perception of Narrative Similarity”. In: *CIKM*. CIKM ’14. ACM, 2014, pp. 321–330.
- [15] Gerald Prince and 遠藤 健一. 物語論辞典. 改訂. 松柏社叢書 言語科学の冒険 4. 松柏社, 2015, p. 259.
- [16] Vladimir Propp. *Morphology of the Folktale*. Vol. 9. University of Texas Press, 2010.
- [17] Vladimir Propp et al. 昔話の形態学. 叢書記号学的実践 10. 白馬書房, 1987, p. 385.
- [18] Alec Radford et al. “Language Models are Unsupervised Multitask Learners”. In: *OpenAI Blog* 1.8 (2019).
- [19] Nils Reiter. “Discovering structural similarities in narrative texts using event alignment algorithms”. PhD thesis. University of Heidelberg, Germany, 2014.
- [20] M P Schraagen. “Folktale similarity based on ontological abstraction”. In: *Global WordNet Conference*. 2016, pp. 352–359.
- [21] Abigail See et al. “Do Massively Pretrained Language Models Make Better Storytellers?” In: *CoNLL Association for Computational Linguistics*, 2019, pp. 843–861.
- [22] S Thompson. *The folktale*. Dryden Press, 1946.
- [23] Hans-Jörg Uther. *The Types of International Folktales: A Classification and Bibliography, based on the system of Antti Aarne and Stith Thompson*. Vol. 1-3. Suomalainen Tiedeakatemia, Academia Scientiarum Fennica, Helsinki, 2004.
- [24] Thomas Wolf et al. “HuggingFace’s Transformers: State-of-the-art Natural Language Processing”. In: *arXiv:1910.03771* (2019).
- [25] 福田アジオ. 日本民俗大辞典. 吉川弘文館, 1999.