

バースト解析を用いた小説テキストの物語理解手法の検討

金子 航太[†]荒川 達也^{††}

† 群馬工業高等専門学校専攻科生産システム工学専攻

†† 群馬工業高等専門学校電子情報工学科

1. はじめに

時系列のテキストデータにおいて、あるキーワードを含む文書が一時的に急激に増加する現象を「バースト」と呼ぶ[1]。これまでにニュース記事や Twitter など、いくつかのメディアを対象にバーストを検知することで、トレンドの動向を分析するという手法が研究されている([2][3]など)。それに対し、本研究ではバースト解析の方法を小説テキストに適用することで、作品内のストーリーや場面の流れを大まかに把握するという手法を提案する。小説テキストに対する「物語理解」の研究は近年盛んに行われているが([4]など)、バースト解析を用いることで比較的手軽に登場人物の活躍度や場面転換等を大づかみに抽出することができ、小説読書支援への応用も視野に入ると考えている。

2. 提案手法

本提案手法では、対象とする小説テキストの各章を時系列文書と考え、指定したキーワードの出現頻度が特定の章において急激に増加する現象を「バースト」と定義する。キーワードとしては人名や地名、アイテムなどさまざまなものが考えられるが、今回の試作システムでは人名(登場人物名)のみ対象とした。各登場人物のバーストを検知することにより、その人物が活躍する章を抽出することが本システムの目的である。バースト解析の手法として本システムでは、Kleinberg のバースト検知アルゴリズム[1]を使用した。

3. 実行結果

現段階ではオルコット作『若草物語』やバーネット作『小公女』、夏目漱石作『坊っちゃん』(すべて青空文庫版)にて実行結果が得られている。実行結果の妥当性を確かめるため、システムにより得られた各登場人物のバースト箇所と、目視により選定した各登場人物の活躍箇所との比較を行った。

以下に『小公女』を使用した実験結果を示す。表 1 にバースト箇所と活躍箇所を並べた表を示す。また、登場人物の章ごとの登場割合(一部抜粋)を図 1 に示す。ここで、「ジュフラアジ」や「バアロウ」といった登場人物は物語全体では滅多に活躍しない脇役であるが、表 1 や図 1 をみると、確かにそれぞれ活躍する章でバーストが検知されているとわかる。次に、各登場人物のバースト箇所と活躍箇所を用いて F 値を計算したところ『小公女』では各登場人物を平均して約 0.89 となった。また、『若草物語』や『坊っちゃん』についても同様な実験を行い F 値の平均を算出したところ、前者は約 0.93 となり、後者は約 0.59 となった。これらは

全体として良好な値であるが、『坊っちゃん』の場合は他の 2 作品と比べ値が低かった。その理由として、一章の分量が比較的多く、活躍シーンをうまく決められなかったという点や、他の他の 2 作品に比べはっきりと人物名を記していないことが多いといったことなどがあげられる。

表 1 実行結果:『小公女』(全 19 章)

キャラクター	バーストした章	実際に活躍した章
ベッキー	5 6 7 15 16	5 6 7 16
アアミンガアド	3 8 9 15	3 8 9 15
ジュフラアジ	2	2
バアロウ	7	7

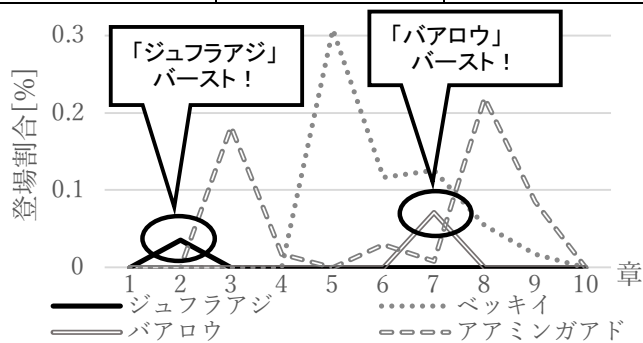


図 1 『小公女』登場人物の登場割合(一部抜粋)

4. まとめ

Kleinberg のバースト検知アルゴリズムを用いて、小説テキストを解析することにより登場人物の活躍シーンを検出する方法を提案した。現在までに 3 つの小説で実験を行い妥当な結果を得た。今後はより多くの小説にて実験を行い、また、それらの小説を読んだ人を対象にアンケートなどを行って実行結果の評価を行っていきたい。

参考文献

- [1] J. Kleinberg, "Bursty and Hierarchical Structure in Streams," in Proceedings of the eighth ACM SIGKDD International Conference on Knowledge Discovery, pp.91-101, 2002
- [2] 高橋佑介ほか, 「ニュースにおけるバースト キーワードの話題への集約」, 第 3 回 DEIM フォーラム論文集, 2011
- [3] 佐々木謙太郎ほか, 「Twitter におけるキーワードの出現周期の自動判別手法に関する検討」, 2013 年度人工知能学会全国大会論文集, 2013
- [4] 山本聖也ほか, 「推理小説の難易度評価のための犯人推定」, 言語処理学会 第 21 回年次大会 発表論文集, 2015