

あらすじに基づくライトノベルの定義作成

大谷紀子 研究室

0632140 外山 雄一

1. 研究の背景と目的

現代日本のサブカルチャー文化の中で生まれたライトノベルと呼ばれる小説がある。ライトノベルは、その他の小説との境界は曖昧である。近年のライトノベルは作品と読者層が多様化している。ライトノベル作家には、ライトノベルに加え一般文芸にも執筆している作家も存在する。また、既存の枠で分類することのできない作家の登場など、今までの概念から大きく広がりを見せている。出版社も明確にライトノベルを謳っているレーベル以外では、ライトノベルとそれ以外の一般文芸の線引きを行って出版しているわけではない。

現在までにライトノベルには様々な定義が提案され議論されている。しかし、提案されたどの定義も例外が存在する。もしくは決定的なものが足りないなど未だに定義は定まっていない。現在までに挙げられたライトノベルの主な定義を以下に示す。

- ①ライトノベルを発行しているレーベルから出版されている。
- ②出版社がライトノベルと宣言している作品である。
- ③マンガ・アニメ調のイラストを多用している。
- ④キャラクターを中心として作られている。
- ⑤青少年あるいは中高生を読者層に想定して執筆されている。

現状では「ライトノベル系レーベルから発売されている、アニメや漫画調のイラストを利用している作品群」という定義でおおむねライトノベルであることが判別できる。近年、ライトノベルの一般文芸化や一般文芸のライトノベル化などが増加し、ライトノベルの一般文芸への越境が増加しているため、現状の定義では徐々にライトノベルの分類が不確かになると予想される。本研究は新たにライトノベルの定義を作成することで、ライトノベルと一般小説の境界を明確にすることを目的とする。

2. ライトノベルの定義の抽出

ライトノベル100冊と一般小説100冊のあらすじを形態素解析する。まず、tf-idf法を用いて抽出した単語の重要度を算出し、重要度の高い単語を有効語とする。tf-idf法はある文章における単語の出現頻度をtf(term frequency)、単語が出現する文章の割合をidf(inverse document frequency)とし、両者の積で文章における単語の重要度を求める手法である。次に小説一つ一つでなく、ライトノベルのカテゴリーとしての特徴を表す有効語を単語間共起に基づく方法で抽出する。ライトノベルの有効語を参照し、ライトノベルに多く現れる単語を抽出する。抽出した単語から重要度を求め、重要度の高い単語を有効語とする。抽出した有効語をもとに有効語ベクトルを作成する。小説を表すベクトルは、当該小説に含まれる有効語ベクトルの和で求められ、カテゴリーのベクトルは各小説のベクトルの平均で求めら

れる。各文章が独立して存在する場合はtf-idf法が有効であるが、文章が属するカテゴリーで分類する場合には単語の共起状況を考慮したベクトル空間が有効である。

ライトノベルの分類には形態素解析システムkncoderで作成されるファイルを使用する。ファイルを本システムに入力し、ライトノベルの判別をする。解析した小説のベクトルを計算し、ライトノベルのベクトルの類似度と閾値の大小関係より判別する。小説とライトノベルの類似度は、小説のベクトルとライトノベルのベクトルとのなす角の余弦値で表す[1]。

3. 評価実験

実験には有効語の抽出に使用した小説とは別にライトノベル100冊、一般文芸100冊を用意し、本システムを通して分類し定義を検証する。評価方法は本システムが出した結果に対しての再現率と適合率を求めることで評価する。

評価実験の結果、ライトノベル100冊のうち73冊が正確に判別され、一般文芸100冊のうち34冊がライトノベルと判別され、正しく判別されたのは66冊であった。全体の再現率は73.0%、適合率が68.2%で、全体の正答率は69%であった。

一般文芸でライトノベルと判断された小説34冊のうち19冊がファンタジー小説であり11冊が恋愛小説であった。今回、評価実験に使用した一般文芸にはファンタジー小説と恋愛小説は、それぞれ30冊含まれているので、ファンタジー小説は半数以上が、恋愛小説は1/3以上が、ライトノベルに分類されたことになる。今後、カテゴリー別に分けた上でのライトノベルと一般文芸の分類を踏まえたうえで定義を作成しなおす必要性があると考えられる。

4. 結果・考察

tf-idf法で抽出した有効語の結果は、全体的に主人公の名前などの人名が重要度の上位に位置することが多いとわかった。一般文芸には人名が一冊中に最大3回出現するのに対して、ライトノベルでは4回～6回出現するものが存在した。ライトノベルのカテゴリーとしては少女、少年、学園、高校生、平凡、普通、世界、ファンタジーなどが上位を占めていた。以上の結果よりライトノベルのあらすじには、全体的なストーリーの説明よりも主人公の特徴や所属、小説内の世界の説明やキャラクター、世界観などの設定を重視する傾向があると推測する。

また、形態素解析において名詞から人名を抽出する方法が不十分である、特に「啓太」という人名が「啓」と「太」に分割され、結果に影響していることがわかった。今後の課題として名詞から人名の抽出する手法の再検討があげられる。

以上により本研究で作成した定義は、再現率と適合率が高い数値を示しておらず不十分であったが、あらすじを基にライトノベルと一般文芸を分類できることがわかった。今後の改善により定義として十分なものができると考える。

参考文献

- [1] 大谷紀子, “情報検索におけるベクトル空間モデルの応用,” 武蔵工業大学環境情報学部紀要, 第五号, pp99-109, 2004.