5S-06

# 取り合わせ語による意外性を創出する俳句の自動生成方式

信末竜空† 秋吉 政徳†
神奈川大学†

## 1 はじめに

日本の伝統的な詩である俳句は世界一短い詩とも言われ、わずか17音という短い音数の制約を持つ。そのような厳しい制約の中で、俳人たちは読んだものに新たな発見を与えたり、感動を与えたりするような様々な情景を巧みに表現する。廣田ら[1]はSeqGANを用いた俳句生成を行なったが、生成の際に単語の意味レベルでの構成と考慮をしておらず、単に俳句と同じような形式の単語列を出力するに留まっている。

俳句は詠み方によって「一物仕立て」の句と「取り合わせ」の句の 2 種類に分類される。「一物仕立て」とはひとつの事物に焦点を当てて詠む手法で、「取り合わせ」とは 2 つの事物の組み合わせによって面白さを演出する手法である。「取り合わせ」では、選んだ事物の組み合わせにより生まれる俳句の意外性が大切であるといわれている。

本稿では、読んだ人の感性を刺激するような面白い俳句を生成するために「取り合わせ」に着目し、これによって俳句としての意外性を創出する方式を提案する。

## 2 提案方式

### 2.1 アプローチ

取り合わせの句では、しばしば一見関係ないと思われる予想外な事物の組み合わせによって、その句の世界観が表現される。一方で俳句としてそれらを詠むとその主題は明瞭であり、そこに意外性を感じる。人が意外性を感じるためには取り合わせの“予想外性”とひとつの俳句としての“主題明瞭性”が必要である。しかし、予想外な取り合わせをするとそれだけはっきりとした主題を読み手に感じさせるのは難しくなるという観点から、この二つはトレードオフの関係にある。そこで、GANの敵対的学習を利用して、予想外性と主題明瞭性のバランスを取り入れた学習方式を提案する。提案方式の構成を図1に示す。基本的な構造はGANと同様であるが、大きな違いはGeneratorが取り合わせ生成部と俳句生成部で構成される点である。取り合わせ生成部はひとつの単語を入力として、入力に合った取り合わせ単語をひとつ選択し、それら二つの単語を一対のペアとし出力とする。俳句生成部は、取り合わせ生成部によって得られた単語ペアを入力として俳句を出力する。

Haiku generation method to inject sence of surprise using GAN with assorted words
†Ryuku Nobusue, Masanori Akiyoshi, Kanagawa University

学習についてはDiscriminatorは判別結果が正しくなる方向に、俳句生成部は生成した俳句がDiscriminatorにサンプルデータであると判断され、さらに主題明瞭性が高くなる方向に、取り合わせ生成部はDiscriminatorにサンプルデータであると判断され、さらに取り合わせが予想外なものとなる方向に行われる。Discriminatorはサンプルデータの俳句か生成された俳句のどちらかを入力として与え、それがサンプルか生成俳句どちらであるかを判別した結果を出力する。

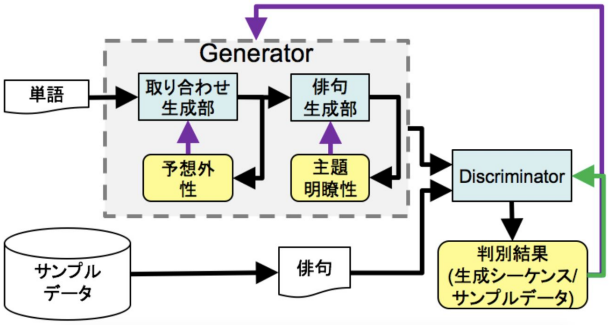


図1:提案方式の構成

### 2.2 トピックモデルによる予想外性と主題明瞭性の算出

トピックモデルのLatent Dirichlet Allocation(LDA)[2]を用いて主題明瞭性と予想外性を算出する。LDAの単語の共起性に基づいた学習から文書をトピックの集まり(トピック分布)の形で表現することができる。主題明瞭性を式(1)により算出した。主題明瞭性は「単語選びに、はっきりとした主題が見られる」と定義する。LDAモデルにより文書のトピック分布を算出した際、主題が明瞭でなければその文書のトピックが分散するという観点から、文書のトピック分布のエントロピーに負の符号をつけることで主題明瞭性を算出している。予想外性を式(2)により算出した。取り合わせ単語の予想外性は「取り合わせた単語が意味的に離れている」と定義する。単語同士のトピック分布が離れているほど予想外性が高くなるという仮定に基づいている。主題明瞭性は俳句が少ないトピックで構成されている時に大きい値をとるのに対し、予想外性はその俳句に含まれる単語のトピック分布が離れているほど大きな値をとる。このような形で俳句の主題明瞭性と取り合わせの予想外性を定義することで、これら2つのトレードオフをモデル化する。



Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.

$$clarity = 1 - \frac{entropy(LDA(d))}{entropy(UD_k)} \quad (1)$$

$$unexpectedness = D_{KL}(LDA(w_1) \| LDA(w_2)) \quad (2)$$

$d$ ： 文書 (俳句)/単語
$k$ ： $LDA$ のトピック数
$entropy(\cdot)$ ： 分布関数のエントロピー
$LDA(d)$ ： $d$ のトピック分布
$UD_k$ ： $k$ 次元の一様分布
$w_1$ ： 入力単語
$w_2$ ： 取り合わせ単語
$D_{KL}(\cdot)$ ： 分布関数の $KL$ ダイバジェンス

## 3 評価実験

学習に利用した俳句は、俳句ポスト365[3]からスクレイピングによって取得した2,449句である。提案手法により生成された俳句の性質を調べるために8名の学生を対象にアンケート調査を実施した。アンケート項目は以下の6つである。

- 次の俳句について、情景としてのイメージが成立するものを全て選んでください。
- 次の俳句について、予想外な展開のものを全て選んでください。
- 次の俳句について、独自性のあるものを全て選んでください。
- 次の俳句について、意外性のあるものを全て選んでください。
- 次の俳句について、人が詠んだものか AI が生成したものか、あなたの考えに最も近いものを選んでください。
- 意外性を評価する際に重要だと考えられる要素

それぞれの質問項目で選択対象の俳句は同一の 31 句である。これらの俳句は、筆者が意外性を感じるか面白いと思った人間の詠んだ俳句 16 句と提案手法により生成された句から主に定型句になっていたもの 12 句、従来手法[1]により生成されたもの 3 句である。

各評価要素と意外性の関係性を調べるため、それぞれを選択した人数の相関係数と、意外性があると選択したときに評価要素を選択している条件付き確率を算出した。その結果を表1に示す。相関係数、条件付き確率のどちらもイメージ成立性が低い値であった。「俳句が意外性あり」と選択した人数について分散分析を行なった結果、提案方式、従来方式、人の詠んだ句の3項目について有意な差は見られなかった。

表 1: 各評価要素と意外性の関係性

| | 予想外な展開 | イメージ成立性 | 独自性 |
|---|---|---|---|
| 相関係数 | 0.271 | −0.190 | 0.369 |
| 条件付き確率 | 0.571 | 0.190 | 0.595 |

## 4 考察

俳句の意外性と評価要素があると選択した人数について相関係数を算出した結果、いずれの評価要素も強い相関は見られなかった。俳句の意外性ありと選択した人数について分散分析や、破調と定型句の意外性についてのt検定でも$t(29) = -0.25, p = 0.40$となり、有意な結果は得られなかった。一方で、個々の俳句について意外性ありと判断した回答者数の分散は大きく、有意な結果などを得るためには回答者数を増やすことが必要であるとも考えられる。また、イメージ成立性については相関係数が負の値をとり、条件付き確率もかなり小さい値となったことから、意外性の評価要素としては適していないという可能性が示唆された。意外性を喚起するような要素についてさらに検討する必要がある。

## 5 おわりに

本稿では、意外性のある俳句の生成を目的としてGANをベースとした俳句生成方式を提案した。Generatorを取り合わせ生成部と俳句生成部に分け、取り合わせ生成部には取り合わせの予想外性を、俳句生成部には主題明瞭性を損失関数に加えることで、これらの相反する目的を持って学習する。これによる取り合わせ生成部と俳句生成部とDiscriminatorの3つのモデルによる敵対的学習の方式を提案した。アンケート調査を行い、提案方式により生成された俳句の特徴を調べたが、統計的に有意な結果は確認できなかった。今後の課題として、今回はSeqGANを基にしたモデルであることから単語を単なるラベルとして扱ったが、単語の分散表現などを用いることで意味的な部分の学習をより工夫することが生成俳句のクオリティ向上に必要であると考えられる。

## 参考文献

[1]廣田敦士,岡夏樹,荒木雅弘,田中一晶,“学習データセットを分けたseqGANによる俳句生成”,言語処理学会第24回年次大会発表論文集, pp.1292-1295(2018)
[2]Blei, David M., Andrew Y. Ng, and Michael I.Jordan.“Latent dirichlet allocation.” Journal of machine Learning research 3.Jan (2003), pp.993-1022(2003)
[3] http://haikutown.jp/post/



Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.