

# 季語情報付与による俳句自動生成器の改良と評価

加藤智一<sup>1</sup> 竹田晃人<sup>1</sup>

<sup>1</sup>茨城大学 大学院理工学研究科

{21nm425n, koujin.takeda.kt}@vc.ibaraki.ac.jp

## 概要

知能情報システムの分野は日進月歩であるが、いまだ人間の「美意識」の表現は苦手としている。これらを言語表現の分野から克服および再現するため、世界で最も短い定型詩である「俳句」と自動文章生成システムを組み合わせた知能情報システム「AI一茶くん」[1][2]が北海道大学を中心に2017年から研究されている。本研究では、知能情報システムに俳句を学習させる際、歳時記および季語辞典をもとに俳句特有の「季語」に関する情報を同時に与えることで、俳句固有の「季語」の観点から俳句の生成精度の向上を図ることを目的とし、俳人の専門的見地から本研究で生成された俳句、人間の俳句、先行研究の俳句の比較検討および再評価を行う。

## 1 はじめに

知能情報システムは、人間の「感性」や「美意識」の表現をいまだ不得意としている。これらを言語表現の分野から克服および再現するため、世界で最も短い定型詩である「俳句」と自動文章生成システムを組み合わせた知能情報システム「AI一茶くん」[1][2]が北海道大学を中心に2017年から研究されている。

俳句は「季語」という固有の概念によって発達した独自の言語表現を特徴とする文学でありながら、世界で最も短い5・7・5の計17音の定型によって知能情報システムでの再現がしやすく、俳人による数値的な評価も容易であることから、知能情報システムの機能実証に最適である。

「AI一茶くん」は、任意の一単語や、画像から抽出した一単語から、ニューラルネットワークの一種であるLSTM (Long Short-Term Memory) を用いて、俳句を自動で生成することができる。この研究は自動文書作成システムの先進的技術開発及び知能情報システムの機能実証を目的とし、人間の俳句と人工知能が生成した俳句の差異から「知能とは何か」「人

と機械のコミュニケーションはどのように行われるべきか」という問題について理解を深める研究である[2]。

しかし先行研究では、生成した俳句から季語を一つ以上含むものを抽出し、残りの俳句を大量に内部で捨てているという問題点がある。このような手法上の問題が先行研究での俳句の生成精度が向上しない要因となっていると考えられる。

本研究では、知能情報システムに俳句を学習させる際、歳時記および季語辞典をもとに俳句特有の「季語」に関する情報を同時に与えることで、俳句固有の「季語」の観点から俳句の生成精度の向上を図ることを目的とし、俳人の専門的見地から本研究で生成された俳句、人間の俳句、先行研究の俳句の比較検討および再評価を行う。

## 2 俳句の自動生成器と学習方法

### 2.1 「俳句」及び「季語」の定義と収集

本研究では、「季語」を一つ以上含み、韻律が5・7・5となる有季定型の俳句を「俳句」と定義し、自由律俳句および無季俳句については考慮しないこととする。本研究では現代俳句協会の「現代俳句データベース」[3]の季語の解釈に則って、季語として現代俳句データベースに立項されている単語を季語とし、傍題は含めないものとする。

季語及び学習の教師となるべき優れた俳句の収集は、現代俳句協会の許可のもと、現代俳句データベースから行った。

### 2.2 動作環境とシステム

システムはpythonで構築するものとし、形態素解析エンジン「janome」[4]を用いて俳句を形態素に分割し、マルコフ連鎖によって表現される離散的なベクトル場の状態を、自然言語処理モデル「word2vec」[5]で学習するものとする。俳句自動生成システムは、「前処理」「学習」「出力」の3段階に分けられる。

「前処理」の段階では、形態素解析エンジンを用いて俳句を形態素ごとに分割する。新しく出現した形態素から順に形態素リストに格納する。そして形態素リストをもとに俳句を単語ごとの ID に変換する。

「学習」の段階では、ベクトルに変換された形態素を入力し、俳句の文脈上その直後にはどのような形態素が出現するかを学習させる。

「出力」の段階では、学習を行ったシステムに任意の一単語を確率的に与える。そして、文脈上つきに出現する確率のある単語から、5音または7音を越えない単語を選択する。5音または7音を満たすまでこの操作を繰り返すことで、5・7・5の俳句が生成される。

本研究では現代俳句データベースをもとに、その単語が季語であるか否かの情報を与える。

## 2.3 手法

人間が恣意的に定義した記号である「言語」をそのまま数値的に扱うことは難しいが、文字や単語をベクトル空間上に埋め込み一つの点として捉える「単語分散表現」を用いることでこの問題は解決できる。言語を数値的に扱えるということは、すなわち学習・解析することが可能となる。word2vecは、この単語分散表現を用いた2層のニューラルネットワークで構成される自然言語処理モデルである。形態素を空間上の点に置き換え、形態素間のつながりをベクトル表現で表し、言語処理を行うことができる。シンプルな構造であるため、大規模な言語データによる分散表現学習が現実的な計算量で可能となるという特徴がある。

マルコフ連鎖は、未来の挙動が現在の値だけで決定され、過去の挙動と無関係であるマルコフ性をもち、とりうる状態が離散的である確率過程を指す。本研究では、形態素を空間上の点に置き換えるため、空間上の各点及びそれらを結ぶベクトルの状態は、各ステップ時間において離散的である。したがってマルコフ連鎖を用いることができる。

本システムの手法は、俳句を形態素ごとに分割し、形態素を空間上に点として配置し、形態素間の共起関係を表現するため、俳句の文脈において連続して出現する形態素（点）どうしを結ぶベクトルを生成する。空間上のベクトル場が離散的に変化することで学習させることができる。

## 3 出力結果

生成した俳句を一部示す。

- ・法華寺の花の世は旅かなしめる
- ・火の山に水の音ある大暑かな
- ・鶏頭のみんだまって水握る

のように、俳句として形になっているものも生成された。しかしこのような日本語の文法の観点から成立している句は少なかった。

- ・て穴子の麦の穂触れや残り鴨
- ・にありしんひては雪に来ん来ぬとる

のように助詞で始まる句や、不自然に助詞が続く句が多数生成された。これは形態素解析エンジンによって俳句を形態素ごとに分割する際に、俳句が現代語でなく古語や文語で表現されるために誤りが生じる点や、付属語で句が始まらないようにする予防をしていなかった点などが原因として考えられる。

## 4 比較検討

本研究で出力された俳句と、先行研究での俳句、人間の俳句を比較検討する。

人間の俳句については、作風の偏りが無いよう全国の俳句誌から俳句を収集している「俳誌のサロン」[6]からランダムに10句収集した。先行研究の俳句については論文から収集する。ただし、「俳誌のサロン」や先行研究の論文に掲載される段階で人間の恣意が介在しており、本研究の俳句の収集についても同様に、日本語として成立しているものの中から10句収集した。したがって、いずれにしても人間の恣意が介在している中での比較検討となる。

本研究には、全国規模で活躍している5人の俳人に協力してもらい、各評価者が計30句に0~10点の点数をつけることとする。その句の「得点」は、各評価者がつけた0~10点の点数の平均値とする。点数の基準については、全国高等学校俳句選手権大会の採点基準[7]に基づき作成した「本研究の採点基準」にしたがって行う。

本研究俳句10句をA群、人間の俳句10句をB群、先行研究の俳句10句をC群とし、それぞれの句に $A_1 \sim A_{10}$ ,  $B_1 \sim B_{10}$ ,  $C_1 \sim C_{10}$ のIDを振る。

次に、A群、B群、C群の句の得点をそれぞれ、 $X_1, \dots, X_{10}, X_{11}, \dots, X_{20}, X_{21}, \dots, X_{30}$ とする。各句群の平均値は次のようになった。

- ・A群の得点平均値 5.28点
- ・B群の得点平均値 5.84点

・ C群の得点平均値 5.46 点

各句群の得点平均値は全て 5 点台であり、比較すると、(人間の俳句の得点) > (先行研究の俳句の得点) > (本研究の俳句の得点) となる。

ここで、ウィルコクソンの順位和検定を用いて、各句群の評価に有意差があるかどうか調べる。本研究では、「本研究の俳句と先行研究の俳句」「本研究と人間の俳句」「先行研究と人間の俳句」3 通りについて、有意水準 5% で検定を行う。

まず、本研究の俳句と先行研究の俳句の比較を行う。

帰無仮説を、「A 群と B 群の得点の母平均に差がある」と定める。A 群, B 群の得点,  $X_1, \dots, X_{10}, X_{11}, \dots, X_{20}$  を並べ、得点の高い順に 1 位から 20 位の順位を割り当てる。なお、得点と同じ場合は平均順位を割り当てる。A 群の句に割り当てられた順位の和を統計量  $T_a$  とする。B に割り当てられた順位の和を統計量  $T_b$  とする。データ数がともに 10 のとき、有意水準 5% で検定すると、 $T_a$  または  $T_b$  のうち、値の小さいほうが 78 以上 132 以下にあるとき帰無仮説が棄却される。

A 群の順位和  $T_a=122.5, T_b=87.5$  より、どちらも 78 以上 132 以下のため帰無仮説は棄却され、A 群と B 群の句の得点の母平均に有意差はなかった。

同様に「人間の俳句と先行研究の俳句」「本研究と先行研究の俳句」の比較を、有意水準 5% で行った。結果を次にまとめる。

・ A-B 群:  $a=122.5, b=87.5$

・ B-C 群:  $b=91.0, c=119.0$

・ A-C 群:  $a=112.5, c=97.5$

したがって、「人間の俳句」「本研究の俳句」「先行研究の俳句」に得点の有意差は無かった。

## 5 考察と展望

本研究の評価者には、無記名の俳句に点数を付ける際に、任意で句評を記入してもらった。

本研究の俳句には、「単語から単語への飛躍が大きい。」「単語と単語の『取り合わせ』が成功していれば評価が高く、失敗していれば意味や景を結ばず評価が低くなる。」という評があった。これは単語ごとに生成するモデルに起因しており、学習データを増やすことで『取り合わせ』の成功率を上げることが今後必要であると感じた。

人間の俳句については、「全体的に読める。」「人間臭い。」「単語同士の関係が意味で結ばれていて、

機械らしい飛躍は減った。」と評があった。

先行研究の俳句には、「連語や成句などによる単語自体の意味の希薄さが強い」という評があった。俳句において一般的に成句や連語が成功するのは難しい。例えば、「犬も歩けば棒に当たる」という成句は日本語として自然であるが、「春の空犬も歩けば棒に当たる」という俳句にしたとき、実際に眼前で犬が棒に当たる景が、成句「犬も歩けば棒に当たる」によって打ち消され、「犬」や「棒」という言葉の意味が希薄になるという問題点が生じるからである。5・7・5 の後半になるほど、全体が日本語として意味が自然になるように、語の飛躍を減らしたり、連語や成句を選択したりしてしまう LSTM モデルの問題点が顕著に出ている。

本研究の俳句と先行研究の俳句は「機械っぽい」という同じ評価をもらったものの、実際は真逆の「機械っぽさ」である。本研究の俳句には、1 単語ずつ逐次的に生成するモデルに起因した「日本語としての一句全体の違和感」という意味での「機械っぽさ」が生じており、先行研究の俳句には、LSTM を用いるモデルに起因した「全体を日本語として意味を自然にしようとしすぎるあまり、単語の意味が希薄になる」という意味での「機械っぽさ」が生じている。これは、それぞれのシステムモデルの問題点が顕著に現れている。

最後に先行研究においては日本語として破綻しているものに対し、人為的にフィルタリングを行っているが、これはシステム上で解決できる問題である。今後の研究では、日本語文法及び俳句の表現上の「型」の観点からシステムの改良を試みたいと考える。

## 謝辞

採点に協力してくださった俳人の先生方をはじめ、研究を支えてくださったすべての方々に厚く御礼申し上げます。

## 参考文献

1. 米田航紀, 横山想一郎, 山下倫央, 川村秀憲: 「LSTM を用いた俳句自動生成器の開発」, Proceedings of the Annual Conference of JSAI JSAI2018(0), 1B2OS11b01-1B2OS11b01, 2018
2. Sapporo AI Lab: 「人工知能による俳句の自動生成に関する実証実験「AI 一茶くん」プロジェクト」, <https://www.s-ail.org/works/aihaiku/>
3. 現代俳句協会: 「現代俳句データベース」,

---

<http://www.haiku-data.jp/kigo.php>

4. 「Janome v0.4 documentation (ja)」,

<https://mocobeta.github.io/janome/>

5. 陳博, 瀧田孝康, 泊大貴, 久永忠範:「オープンデータの RDF 化のための項目名のクラスタを使用し

た述語のサジェストに関する研究」, 情報知識学会誌, Vol.30, No.2, pp.236-241, 2020

6. 「俳誌のサロン」 <http://www.haisi.com/>

7. 俳句甲子園実行委員会「『俳句甲子園』における審査基準 (2018 年度版)