小説の単語分散表現に基づく機械学習評価手法の提案
--オンライン小説投稿サイト「小説家になろう！」作品群を中心に--

国際文化研究専攻（国際日本研究講座）
C2KM1006 王劭宇

- 研究目的

本研究は機械学習によって、小説に対する自動評価の実現可能性を探求する。自動評価を実現するために、作品を文字列から単語分散表現（word distribution repression，または単語埋め込み word embedding，単語ベクトル word vector）に変換することで、物語の展開を意味的な側面から捉え、コンピューターによって分析を行うモデルを開発した。既存した何らかのルールによって評価するシステムを作るという従来の思考と異なり、機械学習は要因と結果の間にある対応関係から、一番合致させる計算ルートを見出す方法である。前者は演繹法に該当し、後者は帰納法に該当する。本研究は作品のテキストを要因とし、人気を結果として捉え、対応関係を再現するようにモデルを算出する。現在、ビジネス小説に対して人力で評価しなければならないにも関わらず、編集者の速度と経験に強く依存し、判断の信憑性も人によってばらつきがある。本研究は現状を突破すべく、客観性を持つ自動評価システムを開発し、さらにシステムの信憑性の実態を把握することを目標とする。

- 先行研究

本研究が関与する分野は、物語論・ビジネス小説における経験論・自動創作の三つである。三つの分野はそれぞれの系譜を持ち、互いに関わりなく分断が存在する。物語論（narratology）は文学批評の分野で利用される分析方法であり、ジュネット[1]（Gérard Genette）より提唱され、バルト（Roland Barthes）などの手を通してようやく完成された。物語論は主に構造主義の系譜を受け継ぎ、時制（temps）・焦点化（focalisation）などの諸技法によって作品の機能的分解を主張する。ビジネス小説における経験論は最初ハリウッドの脚本家たちによって積み上げた経験論であり、のち精神分析学の元型[2]（archetyp）理論と

---

[1] ジュネット＝ジェラール（1985）『物語のディスクール―方法論の試み』花輪光訳、水声社

[2] ユング＝カール（1999）『元型論』林道義訳、紀伊國屋書店

神話学の英雄の旅[3]（hero's journey）理論が導入され、体系化された。ハリウッドでの成功をきっかけに、ビジネス小説の業界にも取り込まれ、ある種の業界基準となった。自動創作は自然言語処理（natural language processing、 NLP）に属する課題であり、近年は機械学習の発展につれ、多くの突破があった。その中に transformer 系モデルの発表によって、自動創作はほぼ実現した。

物語論は分析する一方で、創作することに力にならないし、評価も行われない。ビジネス小説における経験論は経験論の故に、主観性を脱せず、人によって解釈不一致が生じることが多い。自動創作は実用レベルまで至っているが、意図的に創作を行うわけではないため、生成された文章は物語として質にばらつきがあり、利用者によって選別しなければいけない。

本研究は物語論の従来の方法と異なる分析方法を提案することと、ビジネス小説における経験論の主観的側面を一般的・客観的な手法によって補足することと、自動創作の不完全性を自動評価によって補完し、自動創作の質が保証されることが期待される。

- 研究対象

本研究は娯楽小説の代表として、小説投稿サイト「小説家になろう！」[4]に掲載された合計約 95 万本の作品を対象作品群とする。「小説家になろう！」のテキストと閲覧情報はすべてリアルタイムで公開されているため、サンプルサイズも十分大きく、作品群の実態も捉えやすく、テキストと評価の対応関係も分かりやすい。出版品の場合、テキストも発行部数も企業の内部情報に該当するため、取得することが難しい。それに、発行部数の計算と作品の宣伝などの方法も企業によって異なってくる。そのため、本研究は「小説家になろう！」を中心に提案方法を検証し、方法の実行可能性を検討することとする。

- 研究方法

本研究は実験一と実験二によって検証していく。実験一はテキストによって人気を予測する機械学習モデルを開発した。しかし、タスクの前提を正しく設定できてなかったため、正しく予測できず失敗した。実験一は回帰問題としてタスクをとらえたのである。それを踏まえ、実験二は前提を見直して分類問題に設定し、新しく開発したモデルが予測することに成功した。二つの実験は機械学習の手法でモデルを構築することで、コンピュータによる自動評価の可能性を検証する。

本研究は文字列を単語分散表現に変換することで、作品の意味的特徴を分析可能にする。従来の自然言語処理のタスクは文字列のままで処理を行うことが多いが、文章を生成することはできるものの、意味レベルでの解析作業はできない。文字列は記号表現 (signifiant) であり、単語分散表現は記号内容(signifié)である。単語分散表現によって、単語の意味を複

---

[3]キャンベル＝ジョセフ（2015）『千の顔をもつ英雄[新訳版]上』倉田真木訳、早川書房

[4] https://syosetu.com アクセス：2024/1/8

数の特徴量の集合によって表すことができる。物語を時系列として捉えると、文字の順番で並ぶ単語分散表現も各シーンにおける意味として違うシーンを渡る抽象的な雰囲気の変化がコンピューターによって捉えられる。意味的変化のパターンによって作品の人気を予測することが可能である。

単語数 n の作品を次元数 d の単語分散表現に変換すれば、n*d の二次元配列になるので、サイズ n*d の細長い写真に見られる。そのため、本研究は画像処理の方法を作品の解析に活用することになっている。

実験のモデルは機械学習によってテキストと人気を合致させるように最適な計算ルートを見出されたものである。サンプルを再現できれば、サンプルのルールが関数に内包されるので、サンプルにフィッティングできる関数を見出せばいいという思考である。前提が異なるため、実験一は PoolFormer を基礎としてモデルを構築したが、実験二はオートエンコーダ[5]（Auto Encoder、AE）を基礎としてモデルを構築した。前者は人気の度合を直接に予測する一方、後者は人気作品だけでモデルを訓練し、入力が認識されるかどうかによって種類を判断する間接的な方法を採用した。

実験一のモデルは Yu[6]の研究によって PoolFormer が構築される。PoolFormer は transformer の派生型である。transformer は複数層の self attention という文脈各所の互いの相関性を計算する構造でできたオートエンコーダモデルであるが、self attention を pooling という文脈を混ぜるプロセスで取り替え、patch embedding という入力を分割してエリアごとに縮小させる構造を self attention の前に置き、オートエンコーダの encoder 部分だけ残すモデルが PoolFormer である。

実験二のモデルは ResNet[7]（residual network）によってオートエンコーダの構造が構築される。ResNet は畳み込みニューラルネットワーク（Convolutional Neural Network、CNN）という入力を分割し、エリアごとに判断する構造を複数層重ねるストラクチャーである。それを反対方向繋がることでボトルネック状のオートエンコーダが作られる。作品は長さが揃わないため、文脈で計算する複数層の self attention の transformer の encoder 部分がボトルネックに仕込まれる。

---

[5] G. E. Hinton (2006) Reducing the Dimensionality of Data with Neural Networks, SCIENCE vol.313

[6] Weihao Yu etc. (2022) “MetaFormer Is Actually What You Need for Vision”, Conference on Computer Vision and Pattern Recognition, IEEE

[7] Kaiming He, etc. (2015) Deep Residual Learning for Image Recognition, CVPR, the Computer Vision Foundation

実験一は R^2 係数が低下し、有効な予測を出さないと考えられるため、失敗した。実験二は T 検定によって人気作品と他の作品においての差が有意義だと証明されたため、成功した。ただし、実験二のモデルは人気作品に対して 8 割の正解率を達するにも関わらず、その他に対して 5 割の正解率しか出さない。言い換えれば、実験二のモデルはその他の作品を誤って人気作品に判定してしまうという偽陽性の可能性が高いため、人気作品である判定は信憑性が相対的に低い。違うサンプルに対するパフォーマンスの差がこれからの未解決問題になる。

- 結論

実験一だけによって自動評価の可能性を証明できないが、実験一の失敗の上で自動評価テストが回帰問題ではなく分類問題であるという仮説が提出され、実験二によって証明された。言い換えれば、テキストが人気における影響は、連続的ではなく、ありなしだけに限られる。つまり、テキストの質に閾値があり、閾値を越えるかどうかによって人気が影響される仕組みになっていることが発見された。仮説が検証された上、実験一の問題が解消し、自動評価ができるようになった。

本研究によって、自動評価の可能性が確認されたが、正解率について実験二は五割の結論に至ったにも関わらず、これ以上改善できる可能性はあるか確認できず、未来の研究に託す。実験二によって、ボトルネックと閾値を調整することで、正解率が改善されることが確認されたが、最適化された場合、自動評価はどこまで正解率が上げられるか確認できていない。少なくとも、テキストから評価までそれなりの影響力が観測され、自動評価しようと従来の方法で人力評価しようと、推測可能な理論的天井になっていると考えられる。テキストに内包される情報を内部要因とすれば、業界環境と単純なランダム性などの外部要因は 5 割未満を占めると推測される。

モデルについて、実験二で発生する問題はオートエンコーダを採用したことによる問題と思われる。オートエンコーダは大量な正常データの中から僅かな異常データを選び出す使い方が普通であるが、本研究はその関係を大量な異常データから正常データを選び出すという逆転する形にせざるを得なかったため、異常データ、つまり人気のない作品を完璧に対応できない。さらに異常データが多数を占めるため、問題が拡大する。その影響によって、実験二は人気のない作品の判定は信憑性が高いが、人気作品の判定は実に人気ではない可能性が混ざっており、信憑性が中途半端になっている。それにしても、予想より判定標準が緩いだけで、作品に対して自動評価を行う目標は達成されたと考えられる。

まとめて言うと、本研究によって分かったことは 3 点がある。まず、テキストを含める内部要因が人気に与える影響は連続的ではなく閾値が存在する。そして、テキストによって作品の人気を推測することは可能であるが、テキストだけで予測できない部分は 5 割未満のほどが存在する。最後、文字列を単語分散表現に変換することで、小説作品に対して機械学習モデルによる自動評価を行うことが可能である。