

言語生成モデルを用いた 実世界情報ホラーストーリー生成システム

高木 亜蘭¹ 濱川 礼²

概要: 本論文では実世界の情報を基にし、言語生成モデルを用いてホラーストーリーを生成するシステムを提案する。恐怖とは人間の生存本能に起因し、危機から回避するために生じる感情であるが、自身の安全が保障された環境下で与えられると快感を得ることが出来る。そのため、今日では映画やゲーム、小説などによって手軽に恐怖を得られるようになり、多くのホラーコンテンツが世に出されてきた。ホラーコンテンツには映像表現や音響効果、記述表現など様々な要素により恐怖を与えているが、多くのホラーコンテンツに共通している重要な要素はストーリー（コンテンツの大枠や背景、流れなどの説明）である。ホラーコンテンツのストーリーは現実と虚構が織り交ぜられており、虚構における非現実的体験により恐怖を得ることが出来る。しかし、現実と乖離しすぎてしまうとその場面のイメージがしにくくなるため、リアリティ（現実味）も重要となる。つまり非現実的の中に如何にリアリティのあるストーリーで恐怖を与えることが課題になる。しかし、既成のホラーストーリーでは人ごとの現在いる環境におけるリアリティを効果的に表現するのは極めて難しい。そこで、実世界情報をストーリー内に組み込み、実世界が舞台となったリアリティのあるホラーストーリーを作成すれば、より恐怖を楽しめるのではと考えた。本研究では実世界情報を撮影した画像内の物体に関する説明文を基に言語生成モデルにてホラーストーリーを生成する。言語生成モデルとしては GPT-2 を用いる。大規模ラベルなしコーパスによって学習された事前学習モデルに対して、ホラージャンルに属する小説群をラベルなしコーパスとして *Fine Turning* することによりホラーストーリーを生成させる GPT-2 による言語生成モデルを作成した。

1. はじめに

恐怖とは捕食者から逃避行動を動機づけるような、生命的危険を予測する判断機能とされている [1]。これは人間の生存本能に起因しており、自分では対処できない事象から「逃げ出したい気持ち」こそが恐怖感情である [2]。そのため、恐怖はマイナスの感情であり、多くの場面では忌避されることが望ましい。しかし、時として恐怖はプラスに働き、快感を得ることが出来る。それは、自身の安全が保障された環境下において恐怖を与えられる時である。恐怖を与えられると脳内ではホルモンの一種であるドーパミン（快樂物質とも呼ばれ、楽しことや幸せなこと、快感を得ることで放出される）が放出され [3]、一時的に快感を得ることが出来る。しかし、多くの場合恐怖を与えられるとその対象から逃避することを念頭に置くため快感を感じにくくなる。一方で、逃避対象を排除した（安全が保障された）環境下では快感を感じることが出来る。安全が保障された環境下で恐怖を得ると脳が危険ではないことを理解してい

るため、逃避という行動をとる必要がなくなり、大いに快感を得ることが出来ためである [2][4]。また、恐怖を体験しながらもその事象を切り抜けることでも快感を得ることが出来る [5]。恐怖という危機的事象に対し、安全が保障された環境下では無傷で乗り越えられるため、達成感を得られると同時に自尊心を高めることが出来る。そのため、手軽にかつ安全な状況下で恐怖を得ることが出来る映画やゲーム、小説などのホラーコンテンツは昨今においても人気を博している（ホラーゲームでは「バイオハザード」[6] シリーズが約 1 億 700 万本の売り上げ [7]、映画では「IT/イット “それ”が見えたら、終わり。」[8] の北米での売り上げが約 3 億 2,748 万ドル [9]）。

ホラーコンテンツには映画やゲームなどに用いられる映像表現や怪談やサウンドノベルなどの音響効果、小説や漫画などの記述表現など様々な要素により恐怖を与えているが、多くのホラーコンテンツに共通している要素としてストーリーがある。ストーリーはコンテンツの大枠や背景、流れなどを説明するものであるため、恐怖を与える上で重要な要素である。ホラーコンテンツのストーリーは怪物や幽霊、異常な人間など現実とはかけ離れた非現実的な事象が

¹ 中京大学 工学研究科 情報工学専攻

² 中京大学 工学部

綴られる。これは、人の本能的欲求の中に現実からかけ離れたものを経験したいという思いが根底にある [5]。そのため、非現実的で恐ろしいストーリーだったとしても興味が沸き、満足感を得ることが出来る。しかし、ストーリーの大部分が非現実的(虚構)で良いとなるとそうではない。虚構の部分が多くなると、場面の想像が出来ず、十分に恐怖を与えることが出来なくなってしまう。つまり非現実的の中に如何にリアリティのあるストーリーで恐怖を与えることが課題になる。しかし、既成のホラーストーリーでは人ごとの現在いる環境におけるリアリティを効果的に表現するのは極めて難しいため、リアリティの程度はすでに収録された文章の精巧性によるものとなってしまう。

そこで、リアリティのあるホラーストーリーを作成するには実世界情報を活用することで実現できるのではと考えた。実世界情報を活用することで、舞台が実世界となったホラーストーリーを形作ることが出来る。これにより、ホラー要素が実世界に即して綴られ、リアリティのある演出をすることが出来るため、より恐怖を楽しめるのではと考えた。本論文では現実世界からの情報を基にホラーストーリーを生成するシステムを開発した。これにより、現実世界の状況に則し、リアリティを演出しつつも恐怖を与えるストーリーを提供する。

2. 関連研究

「ホラーストーリー生成」と「画像からのストーリー生成」の関連研究について記述する。

2.1 ホラーストーリー生成

ホラーストーリーの生成に関する取り組みとして「KATARIBE ちゃん」[10]がある。これは、毎夜 24 時に怪談を生成する twitter の怪談自動生成 bot である。また、この bot を原作とした漫画「カタリベ 人工知能がつぶやく呪いの予言」[11]があり、これは「KATARIBE ちゃん」が生成した文章を基にして、漫画に起こすプロジェクトである。また、ホラーストーリー生成に関して“Shelley”[12]がある。これは、twitter の怪談自動生成 bot であり、掲示板サイト“reddit”にて掲載されている 14 万ものホラー小説を AI に学習させている。自発的にストーリーを生成する他に、他ユーザからのツイートからリレー形式に文章を紡ぐこともできる。これらは、ホラーストーリーを生成することが出来るが、文章のみの入力にしか対応していない。本システムでは画像を入力にすることにより、現実世界とリンクさせたホラーストーリーを生成し、リアリティのある恐怖を与えることが出来る。

2.2 画像からのストーリー生成

画像からストーリーを生成する研究の代表例として Yu らの研究 [14] と、Jing らの研究 [15] がある。これらは、

順序付けされた複数の写真画像を入力として、画像毎でキャプション(説明文)を生成する。その後、写真同士の順序関係を考慮し、文が繋がるよう補完を行い、1 つのストーリーを生成する研究である。Yu らは *Bidirectional Attention-based Recurrent Neural Network (BARNN)* を用いることにより、以前の写真での事象を引き合いに出し、文が繋がるようなストーリーを生成している。Jing らは敵対的トレーニング手法を用いることにより、感情と表現豊かな文体でのストーリー生成をしている。これらの研究では、複数の写真画像から順序を考慮したストーリーを生成することが出来るが、画像のキャプション内容に関係のない文の挿入(画像のキャプション同士を繋げることしかできない)や特定のジャンルに寄せるようなストーリーの生成に対応していない。本研究では、画像から説明文を導き出し、画像の説明内容以外の文を付け加えた特定のジャンルのストーリーを生成する。

3. 提案手法

本研究では、ホラーストーリーのリアリティ向上を図るために実世界情報をストーリー生成の際に取り込むことにした。実世界情報は日々暮らしている世界においての生活状況や環境情報を表しているため、リアルな情報を得られる。このようにして、実世界情報を取り込むことが出来れば、ユーザの生活空間を舞台に設定することが出来るため、リアリティが増すと考えた。

実世界情報は画像から得ることとする。これは、ユーザが手軽に実世界の場面を切り取ることができ、実世界の状況を多く含んだ情報源であるためである。画像に映し出されている情報を基にしてホラーストーリーを生成することにより、実世界での状況を加味したリアリティのあるストーリーが実現できる。生活している身近な環境における怖い出来事が綴られるため、実際に現実で起こりうるのではないかと思わせられ、リアリティ向上に繋がる。画像から実世界情報を取り出すためには、画像に関する説明をテキスト形式で生成する。これは、画像に関する情報を完結的かつ容易な形式で表現でき、またストーリー生成の際に用いる言語生成モデルに対して情報の受け渡しが容易になるためである。説明文の生成には画像キャプション技術を用いる。画像キャプション技術は *CNN(Convolutional neural network)* と *RNN(Recurrent Neural Network)* を用いることにより画像の内容を説明する短い文章を生成する手法である。これにより、画像内の状況を容易に引き出すことが出来る。実世界情報を表す説明文を言語生成モデルへ渡すことにより、説明文を基として文章が生成されるため、実世界を考慮したリアリティのある文章が実現できる。

ストーリーの生成には言語生成モデルを用いる。言語生成モデルとは入力された文に対して次の単語の予測し、テキストを生成する言語モデルである。そのため、開始文を

入力することで文章が自動的に生成されるようなストーリー生成に適したモデルである。本研究では、生成する文章はホラーストーリーであるため言語生成モデルをホラーストーリーの生成に重きを置いたモデルに変更する。このように、言語モデルに対してホラーのジャンルに該当する文章を多量に学習させることにより、ホラーテイストな文章を生成するモデルが獲得できる。

実世界を映した画像とホラーストーリーを生成する言語生成モデルをテキスト形式による説明文によって組み合わせることにより、リアリティのあるホラーストーリーが生成でき、ユーザを楽しませることが出来る。

4. システム構成

本システムの構成図を図1に示す。本システムでは入力画像から画像内の状況を説明する文を生成する「画像説明部」と、説明文からホラーストーリーを生成する「ストーリー生成部」の二つの構成となっている。なお、主な開発言語として *Python* を用いた。

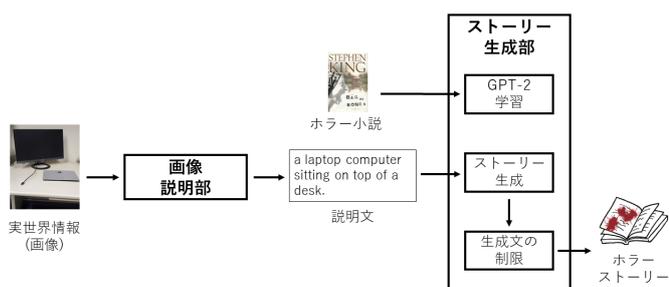


図1 システム構成図

4.1 画像説明部

画像説明部では、入力された画像に対し映り込んだ物体に関する説明文を生成する。

説明文の生成の手法としては“Show, Attend and Tell: Neural Image Caption Generation with Visual Attention” [16] を参考に実装した。これは *CNN* を使用して画像から視覚的特徴を抽出し、*LSTM* (*Long Short Term Memory*) を使用してこれらの特徴を文にデコードする手法である。品質を向上させるために、*Attention* メカニズム [17] が組み込まれている。

使用する認識モデルとしては“MS COCO 2014” [18] データセットで学習された事前トレーニング済みモデル [19] を使用した。事前トレーニング済みモデルの *BLEU* (*Bilingual Evaluation Understudy*) スコアを以下に示す。

- *BLEU-1* = 70.3 %
- *BLEU-2* = 53.6 %
- *BLEU-3* = 39.8 %
- *BLEU-4* = 29.5 %

事前トレーニングモデルを用いた説明文付加結果の例を表1に示す。

表1 説明文付加結果の例

a laptop computer sitting on top of a desk	a room filled with furniture and a book shelf.
	

4.2 ストーリー生成部

ストーリー生成部では画像説明部にて生成された説明文を入力としてホラーストーリーの生成を行う。ストーリーを生成する言語生成モデルとしては *GPT-2* [20] を使用した。*GPT-2* は *Transformer* [21] をベースとした言語生成モデルであり、*Attention* メカニズム [17] が組み込まれている。これにより、高い精度での文章生成が可能となっている。

4.2.1 コーパス

ホラーストーリーの学習用のコーパスには“*All Novel*” [22] に掲載されているホラーのジャンルに属する小説を約 1.5GB 分 (283 冊) を素材して収集した。この素材を小説毎で1行に記し、先頭 (*beginning of sentence*) に<BOS>, 後尾 (*end of sentence*) に<EOS>のタグを付ける。これにより、学習の際に小説毎で区切って学習するラベル無しコーパスができる。

コーパスの形式

```
<BOS>小説 1<EOS>
<BOS>小説 2<EOS>
<BOS>小説 3<EOS>
...
```

4.2.2 GPT-2 による学習

GPT-2 に対して 4.2.1 項で作成したラベルなしコーパスを用いて *Fine Turning* させることにより、ホラーストーリーを生成する言語生成モデルが作成できる。

自然言語処理関連のモデルが公開されているサイト“*hugging face*” [23] より、事前学習済みモデル“*gpt2*”をベースモデルとして *Fine Turning* を行った。“*gpt2*”は *Open AI* が約 800 万の *Web* ページ (計約 40GB) のテキストデータをラベルなしコーパスとして収集し、48 層のネットワークでおおよそ 15 億個のパラメータにて学習した事前学習モデルである。

4.2.3 ストーリー生成

ストーリー生成には開始タグである<BOS>と4.1項にて画像から生成された左の説明文の例を連結させた文を開始文として4.2.2項にて生成された言語生成モデルに入力し、ストーリーの生成を行う。今回は評価のしやすさを重視するため、入力文を除いて計5文の短文を生成するよう設定した。生成結果を以下に示す。開始文は表1にて生成された“A laptop computer sitting on top of a desk.”を用いる。また、DeepL[24]による和訳文も示す。

生成したホラーストーリー

<BOS>A laptop computer sitting on top of a desk. An orb floating above it was about to be fired and crashed across the room, in shock but not minding its own business as most of us did when it suddenly hit another person or something. The two guys looked at each other after she had gotten into her chair again (she's always done weird stuff, really) and said that perhaps "it" would help them out any time they got back home to **California**. Then he turned his head over slowly toward me wondering if I could talk with the one who had been watching him and my father from up close for weeks all this week before we finally came along where people tend very differently sometimes than others! So it kind of bothered my dad a little bit more than anything else - even though whatever... Nothing scary has ever happened to someone like how my younger brother usually works things out anyway; except now as he talked about an older woman trying desperately hard while our dad continued their conversation there seemed some sort thereof? What the hell was there until now since nothing has stopped it going around the house so easily without warning?

DeepLによる和訳文

<BOS>机の上に置かれたノートパソコン。その上に浮かんでいたオーブが発射されようとして部屋中に墜落した。ショックを受けながらも、他の人に突然ぶつかったりしても、ほとんどの人がそうであるように、気にも留めない。彼女が再び椅子に座った後、二人は顔を見合わせて（彼女はいつも変なことをしている、本当に）、もしかしたら「それ」はカリフォルニアに帰ればいつでも助けてくれるかもしれないと言った。そして、父はゆっくりと私の方に顔を向けて、自分と父を何週間も間近で見ていた人と話ができるのだろうか、と考えていました。だから、父は何よりも少し気になった。弟のような人には怖いことは何も起こらないのですが、年配の女性が必死に頑張っているという話をしながら、父が会話を続けていると、何かがあるように思えたのです。何の前触れもなく簡単に家の中を飛び回るのを止められなかったのに、今まで一体何があったのだろうか？

4.2.4 生成文の制限

4.2.3項にてホラーストーリーを生成した。しかし、生成文章内に固有名詞(太字で記している)が見受けられる。固有名詞があることにより、画像内の状況と生成ストーリーの間に明確な違いが生じてしまい、違和感が生じてしまう。そのため、生成する文に対し、固有名詞を含めないよう制限することにより、違和感の解消を図った。

固有名詞を含めないストーリー生成手法としては、生成時に複数文をまとめて生成するのではなく、1文生成した際に固有名詞の有無を判定させるよう文の制限を行った。固有名詞の判定には形態素解析ライブラリであるNLTK(Natural Language Toolkit)を用いた。

以下、制限手法である。

- (1) 開始文を入力する
- (2) 開始文に対する1文の生成を行う
- (3) 生成した1文内に固有名詞があるか判定

ある

生成した1文を破棄し、(1)に戻り再生成

ない

生成した文を開始文の後尾に結合させ、(1)に戻り新しい文を生成

固有名詞を制限して生成したホラーストーリーを以下に示す。開始文は表1にて生成された“A laptop computer sitting on top of a desk.”また、DeepL[24]による和訳文も示す。

文生成の制限を行ったホラーストーリー

<BOS >A laptop computer sitting on top of a desk. This would sound stupid if you're not paying attention and there's the creepy bit - you actually hear things like this occasionally, sometimes as in my house they start to move around etc. If it didn't matter where or when I was at that particular point. It just continued with what normally followed each time its done. After about 3 more hours no activity whatsoever so we could all leave. As usual for me the first morning came early after 3am so my cat had run upstairs from his room one minute late (I left him alone so he wouldn't be upset) but then he started again later coming back up through the main window another 10 minutes earlier looking bored out here which wasn't very nice I guess because our windows were really dark so light meant something needed moving on their own.

DeepLによる和訳文(固有名詞の制限あり)

<BOS >机の上に置かれたノートパソコン。注意して見ていないとバカバカしく聞こえるだろうし、不気味なところもある。実際にこのような音が時々聞こえてくるし、私の家のように彼らが動き出したりすることもある。もし、私がある時点でどこにしようと、いついようと、それは問題ではありませんでした。それはただ、いつものように続いていました。さらに3時間ほど経っても何の動きもないので、私たちは全員帰ることができました。いつものように、最初の朝は3時過ぎに来たので、私の猫は1分遅れで自分の部屋から2階に駆け上がったのですが(怒られないように放っておいた)、その後、10分前に再びメインの窓から上がってきて、ここで退屈そうにしていました。

5. ホラーストーリー生成結果

4.2.4項の文生成の制限を行ったホラーストーリーからは固有名詞の記述が見られなかった。そのため、ストーリーの内容に違和感がなくなり、リアリティの向上に繋がったと考える。

4.2.3項の“An orb floating ~ in shock”や4.2.4項の“This would sound ~ move around etc.”のようにホラーストーリーを彷彿とさせるような文の生成が来ている。また、“A laptop computer sitting on top of a desk.”の入力文に対し、4.2.3項の“An orb floating above it”のように実世界情報である入力文と整合性が取れた文が生成されている。これにより、実世界を舞台としたホラーストーリーの生成に成功していると感じる。

しかし、4.2.4項の“After about 3 ~ on their own.”のように文単位での生成精度は高いと感じるが、文同士のつながりが取れていないように見て取れる箇所が複数存在する。また、ストーリーの後半に差し掛かるにつれて実世界情報と関連性のない文が生成されていると感じる。これは、実世界情報に先頭にしか入力されていないため、文が生成される内に関連のない文になったと考える。

以上のことにより、本研究では実世界情報からホラーストーリーの生成に成功しているが、文章全体の意味合いや実世界情報の活用の仕方など改善すべき点があると考えられる。

6. まとめ

本システムでは実世界の情報(画像)を基にして、ホラーストーリーを生成した。これにより、リアリティのあるホラーストーリーの生成を実現できた。しかし、いくつかの課題がある。

- 学習ネットワークモデルの見直し

現在、画像の説明文付加とストーリー生成の言語生成モデルが別々になっている。そのため、説明文とストーリーとの関連が不十分であると考えられる。説明文付加とストーリー生成の両者を組み込んだネットワークモデルを検討することで、ストーリーの後半に差し掛かって一貫して実世界情報について記述している文の生成を実現する。

- コーパスの見直し

今回、コーパスにホラーのジャンルに属する小説の全文を用いた。しかし、小説内にはさほど恐怖を感じさせない文表現が多く含まれていると考える。そのため、長文小説を学習させるのではなく、短文もしくは概要分を学習することにより恐怖を与えることが出来る文が収録される割合が多くなり、ホラーテイストが濃いコーパスを作成できると考える。

- 怖さの指標の決定

作成した文が怖いかどうかの定量的指標の決定を行う。指標を設けることにより、ストーリーを生成する際に文の制限をかけることが出来、より怖い文を生成することが出来ると考える。現在、短文小説に対するレビューを指標の検討に入れている。

- 文同士の意味のつながり

文単体での生成精度は高いが、現状では文の前後での意味のつながりが取れていない。そのため、文同士での意味のつながりを定量的に測定することが出来れば、文脈を通して意味が通じる文が作成できると考える。

- 評価

本論文では生成文章の評価は主観評価で行った。しかし、主観評価のみでは正確な評価を下すことが出来ないと考えられる。そのため、文章の生成精度に関して指標

を用いた定量評価や被験者を募って生成されたストーリーに恐怖を感じたか否かのアンケート調査などを行う。

これらを改善することで、より恐怖を与えられるストーリーが生成できると考えられる。

参考文献

- [1] 戸田正直「感情-人を動かしている適応プログラム」 東京大学出版会 (1992)
- [2] 山根一郎「恐怖の現象的心理学」 椋山女学園大学 (2007)
- [3] 米倉 義晴, 菅野 巖, 須原 哲也, 高橋 英彦「感情の中核である扁桃体におけるドーパミンの役割を解明」 (2010)
- [4] Kerry Ressler: Yerkes National Primate Research Center『「恐怖」と「快感」の裏腹な魅力：ホラー作品の人気を脳神経科学と心理学から分析』 WIRED
- [5] Glenn Sparks:Purdue University “Thrill-Seekers Thrive on the Scary” Web MD
- [6] CAPCOM 「バイオハザード」
- [7] CAPCOM シリーズソフト売上 2020 年記録 <<https://www.capcom.co.jp/ir/finance/salesdata.html>>
- [8] New Line Cinema, RatPac-Dune Entertainment「IT/イット “それ” が見えたら、終わり。」
- [9] “Box Office Mojo.” 2017 年記録
- [10] 「KATARIBE ちゃん」 <<https://twitter.com/tsdnst?s=20>>
- [11] 舟「カタリベ 人工知能がつぶやく呪いの予言」 KADOKAWA(2018)
- [12] Massachusetts Institute of Technology: “Shelley” <<https://twitter.com/shelley.ai?s=20>>
- [13] “reddit” <<https://www.reddit.com/>>
- [14] Yu Liu, Jianlong Fu, Tao Mei, Chang Wen Chen “Let Your Photos Talk: Generating Narrative Paragraph for Photo Stream via Bidirectional Attention Recurrent Neural Networks” (AAAI-17)
- [15] Jing Wang, Jianlong Fu, Jinhui Tang, Zechao Li, Tao Mei “Show, Reward and Tell: Automatic Generation of Narrative Paragraph from Photo Stream by Adversarial Training” (AAAI-18)
- [16] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, Yoshua Bengio “Show, Attend and Tell: Neural Image Caption Generation with Visual Attention” (ICML2015).
- [17] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio “Neural Machine Translation by Jointly Learning to Align and Translate” (2014)
- [18] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár “Microsoft COCO: Common Objects in Context” (2015)
- [19] “image_captioning” Guoming Wang
- [20] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever “Language Models are Unsupervised Multitask Learners” (2019)
- [21] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin “Attention Is All You Need” (2017)
- [22] “All Novel” <<https://allnovel.net/>>
- [23] “huggingface” <<https://huggingface.co/>>
- [24] “DeepL” <<https://www.deepl.com/translator>>