

# 双方向 RNN を用いた画像に対する感想コメント生成に関する検討

川西 康友, 出口 大輔, 井手 一郎, 村瀬 洋 (名古屋大学)

A Preliminary Study on Impression Description using Bi-directional RNN  
Yasutomo Kawanishi, Daisuke Deguchi, Ichiro Ide, Hiroshi Murase (Nagoya Univ.)

## 1 まえがき

近年, Twitter や Instagram などインターネット上の SNS では, あるユーザが投稿した写真やビデオなどのマルチメディアコンテンツに対し, 他のユーザがそのコンテンツに関して感想コメントを投稿するなど, ユーザ同士がマルチメディアコンテンツを通してコミュニケーションする機会が増えている。

SNS 上では, ユーザの投稿に対して自動応答する「ボット」と呼ばれる対話システムが多数開発されているが, 既存のボットは, 基本的にはテキストベースの対話を対象としている。こうした対話システムが, 写真やビデオなどのマルチメディアコンテンツを用いた対話をできるようにすれば, ボットと人が自然で楽しい対話をすることができるようになって考えられる。

本研究では, 様々なマルチメディアコンテンツのうち, SNS 上で最も多く扱われている「写真」に注目し, 写真を提示された際に, 人と同じような感想コメントをすることができるような対話システムの実現を目指す。

本発表では, 画像認識に広く用いられる Convolutional Neural Network (CNN) と, 高精度な予測ができると言われている双方向 Recurrent Neural Network (RNN) を組み合わせたモデルを用いて, 画像から感想コメントを生成するシステムの検討をする。

## 2 関連研究

画像認識の分野において, 画像中の物体を認識してその名前を答えるだけでなく, その画像が何であるかを文章の形で答える技術は, キャプションと呼ばれ, 広く研究されている [1]。従来は, 画像に応じて, 適する文を幾つかの候補の中から選択するアプローチがとられていたが, 近年は言語モデルを用いて文を生成するアプローチも多い。最近, 深層学習, 特に RNN の発展により, キャプションにも深層学習が利用されるようになり, 高品質な説明文生成が可能となっている [2]。

## 3 提案手法

従来のキャプションと同様, 本研究で対象とする感想コメントを生成するためには, 画像及び, その画像に対する感想コメントの組が大量に必要である。こうしたデータを収集するため, 本研究では, Twitter の画像付き Tweet に着目する。画像付き Tweet に対する Reply は, その画像に対する感想コメントであることが多いと仮定し, Twitter に投稿された画像付き Tweet と, その Reply の組を収集する。

提案するコメント文生成システムの深層学習モデルを図 1 に示す。このモデルでは, まず ImageNet の画像で学習した VGG16 モデルを用いて画像から特徴抽出を行い, その結果を双方向 RNN

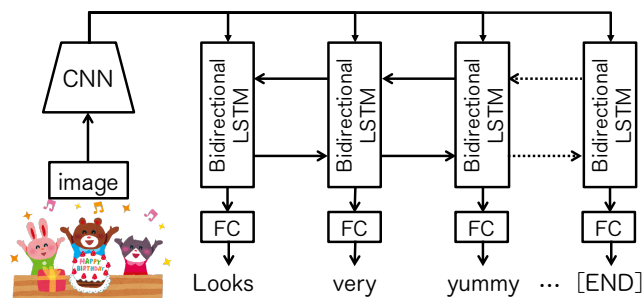


Fig.1 A bidirectional RNN model.

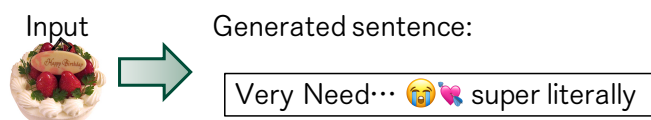


Fig.2 An example of the results.

へと入力する。双方向 RNN の出力をそれぞれ全結合層に入力し, 最終的な単語の尤度分布を得る。その尤度に応じて単語をサンプリングすることで, 最終的な出力とする。

## 4 実験

Twitter から, 約 4000 件の「birthday」「cake」を含む画像付き Tweet を抽出し, それに対する Reply と共に保存し, データセットとした。このデータセットを用い, 前述のモデルを学習した。この時, 文の最大系列長を 15 単語とした。

学習したモデルを用い, 別途収集した画像に対して感想コメントを生成した。結果の例を図 2 に示す。

## 5 まとめ

本稿では, 画像を入力として, その画像に対する感想コメントを自動生成することを目的とし, 双方向 RNN を用いたシステムの検討を行なった。現状では生成される文が不自然な場合があるため, モデルのさらなる改良が必要である。

謝辞: 本研究の一部は科学研究費補助金および, 公益財団法人堀科学芸術振興財団研究助成による。

## 文献

- (1) J. Y. Pan et al., Proc. of ICME, Vol.3, pp.1987-1990 (2004).
- (2) O. Vinyals et al., IEEE T.PAMI, Vol.39, No.4, pp.652-663 (2017).
- (3) K. Simonyan et al., arXiv:1409.1556