

Twitter 上に投稿された画像への深層学習の適応および内容の分析

新田 大悟[†] 山下 晃弘[†][†] 東京工業高等専門学校 情報工学科

1. はじめに

近年, SNS 利用者の増加と共に, SNS 上で個人に非難が集中する「炎上」問題が増加している. 特に, 短文投稿 SNS である Twitter ではその問題が顕著であり, メディアで報道される事例も少なくない. このような炎上問題は時折, 身元の特定に繋がる恐れがあり, 一度特定されてしまうと社会生活が脅かされる他, 利用者に精神的負担を与えるなどの問題が発生する. 炎上の原因として, 投稿者が意図しない大規模な拡散や, 投稿に含まれる画像・文章に問題がある場合が多い. しかし, Twitter に投稿された文章分析の研究[1]はされているものの, 画像を含めた分析はほとんど行われていない. そこで本稿では, ツイートされた画像と文章の特徴量から, 投稿を拡散する Twitter の機能として存在する「リツイート」数がどのくらい集まるかを予測するモデルを構築する. また今回は画像解析を実施するため, 画像付きツイートのみを調査の対象とする.

2. 研究の流れ

2-1. 文章のベクトル化

ツイートから文章を取り出し, トレーニング済みの Doc2Vec[2]モデルを使用して 1,000 次元の特徴ベクトルに変換する. Doc2Vec モデルのトレーニングには Twitter からランダムに集めた 1,544,127 件のツイートを用いた. 学習の際には形態素解析を行い, ベクトルが各品詞の活用形に影響されないように品詞を原形に変換した.

2-2. 画像の前処理

ツイートに含まれている画像を, RGB の 256×256 サイズに変換し, 3×256×256 の配列に変換する. リサイズのアプローチには Bilinear 法を使用した. また, 収集した全画像の画素値の平均値である平均画像を作成し, 各画像から引くことで正規化を行った.

2-3. ニューラルネットワークによる学習

以上の処理によって得られた文章の 1000 次元ベクトルと画像の 3×256×256 次元ベクトルを入力とし, 図 1 に示すニューラルネットワークを構成した. このニューラルネットワークにはユニット間の結合が疎な「畳み込み層」や位置感度を低下させる「プーリング層」など, 画像の特徴量抽出を担う層が含まれている[3].

ニューラルネットワークの構成には Chainer[4]を用いた. また, 画像から特徴量を抽出する畳み込み層は AlexNet[5]と GoogleNet[6]の 2 通りのモデルをベースに構成し, 精度を比較した. これらのモデルは, AlexNet が図 1 における loss1 層のみで誤差を計算しているのに対し, GoogleNet は loss1 層に加えて畳み込み層と連結している loss2 層, loss3 層を使用して誤差を計算している等の違いが挙げられる. GoogleNet が最終的に伝搬する誤差は(1)に示す式によって計算される.

$$Loss = loss1 + 0.3 \times (loss2 + loss3) \quad (1)$$

また, 学習には正解データとして RT の規模を表 1 に基づいて 4 つのラベルに分類したものを用いた. 表 1 中の σ は, アカウントごとのリツイート数の標準偏差とする.

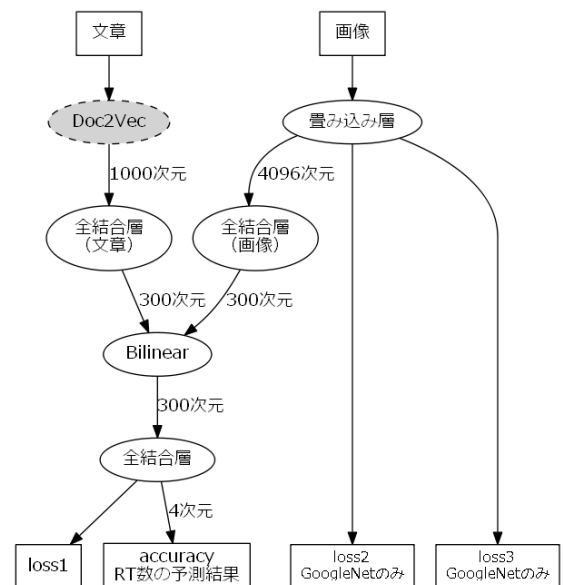


図 1. 使用したニューラルネットワークのモデル

表 1. ラベル内訳

ラベル	RT 数 (σ :RT 数の標準偏差)	データ数
0	0	21256
1	1 以上 1 σ 未満	5415
2	1 σ 以上 2 σ 未満	1811
3	2 σ 以上	2238

3. 実験

今回の実験を行うための訓練・テストデータとして、Twitter からランダムに集めたユーザ 5959 名から直近 3000 ツイートを収集し、その中で画像付きのツイートのみを絞り込んだものを使用した。取得した画像付きツイートは合計 649,369 件となった。また、各ユーザの上位 3,000 ツイートに対し、リツイート数の標準偏差を求め、記録した。学習時間の都合上、実験には収集したデータのうち、30,720 件の画像付きツイートのデータを用いた。収集したツイートは表 1 に示されるように RT 数によってデータ数が偏っている不均衡データであったため、学習の際には全てのラベルのデータ数を最もデータ数が少ないラベル 2 に合わせ、 $1,811 \times 4 = 7,244$ 件を全データ数として整形した。

学習を行う際には 7,244 件のデータを訓練用とテスト用に 7:3 の比率で分割し、学習と精度検証を行った。最終的な結果を表 2, 3 に示す。なお、表 2 で示す Epoch は学習の繰り返し回数、Loss は訓練データと正解データの交差エントロピーを示す。

表 2. 各モデルの Epoch, Loss

	AlexNetBased	GoogleNetBased
Epoch	25	27
Loss	0.064	0.969

表 3. 各モデルによるテストデータの推定精度

ラベル	AlexNetBased	GoogleNetBased
0	47.50%	53.74%
1	41.22%	40.81%
2	43.60%	39.21%
3	45.73%	36.50%

4. 実験結果考察と今後の課題

表 3 より、全ラベルの推定精度の平均値は、AlexNetBased モデルが 44.51%, GoogleNetBased モデルが 42.57% となり、AlexNetBased モデルのほうが僅かながら高い精度を出していることが分かる。しかし、AlexNetBase モデルは Loss が 0.064 と学習がかなり進んでいるのに対して GoogleNetBased モデルは Loss が 0.969 と AlexnetBased モデルほどは学習が進んでいない。しかしながら推定精度の差が 2% ほどであることから、GoogleNetBased モデルは AlexNetBased モデルと比べて表現力が高いと考えられる。これは、GoogleNet が画像の特徴量抽出として畳み込み層に加えて「Inception Module」と呼ばれる層を採用しているからだと考えられる。Inception Module は入力配列に 4 つの異なる畳み込み層・プーリング層を適用し、それらの出力を連結することで表現力を向上させる技法である。実験結果からこの層が有効であることが分かる。

今回の実験では学習時間の都合上、収集したツイートデータ 649,369 件のうちの 5% ほどである 30,720 件のみの使用となった。また、学習の繰り返し回数も 25~27 程である。今後の課題としては、収集した全てのデータを用いた精度の検証や、学習の回数を増やすことを考えている。

5. まとめ

本稿ではツイッター上に投稿される画像付きツイートがどのくらいのリツイート数を集めるかを予測するためのニューラルネットワークを作成し、評価を行った。予測精度は 42~44% となった。

6. 謝辞

本研究の一部は、JSPS 科研費 15K16092 の助成を受けたものです。

参考文献

- [1] 山下晃弘, 上村卓史, 川村秀憲, 鈴木恵二: SNS プライバシー保護とリスク管理の検討, 情報処理学会デジタルプラクティ, Vol.6, No.2, pp.150-158 (2015)
- [2] Radim Rehurek. Gensim Doc2Vec: <https://radimrehurek.com/gensim/models/doc2vec.html>
- [3] 谷岡貴之, 深層学習 (機械学習プロフェッショナルシリーズ) (2015)
- [4] Preferred Networks Chainer: <http://chainer.org/>
- [5] Alex Krizhevsky, et al: ImageNet Classification with Deep Convolutional Neural Networks
- [6] Christian Szegedy, et al. Google Inc: Going deeper with convolutions