

ニューラルネットワークを用いた 国際的なニューストピックスへの国民感情を踏まえた意見表明

樋口廉^{†1} 川野邊誠^{†1}

人がある国際的なニューストピックスに対して何らかの意見を出さずには、その意見には、その人が育った国や文化圏の影響から差異が生じるのが一般的である。現状の AI は、このような差異を考慮した意思や感情を出力に反映するまでに至っていない。そのため、既存の AI による出力は画一的なものとなっている。

そこで、本研究では、国民感情によって受け止め方に差が出るようなニューストピックスに対して生じる意思や感情を出力に反映できる AI の構築を目的としている。そのための一歩として、本論文では、国際的なニューストピックスとそれに向けられた意思や感情の関連性を学習したニューラルネットワークの構築を試みた。その結果、入力した国際的なニューストピックスに対して、三カ国それぞれの立場から、「良いニュース」、「悪いニュース」、「どちらでもない」の3パターンを判断することができたことを報告する。

1. はじめに

人がある国際的なニューストピックスに対して何らかの意見を出さずには、その意見は、その人の意思や感情などの要因から大きな影響を受ける。これらの要因は、その人が育った国や文化圏の影響から差異が生じるのが一般的である。AI における既存の研究では、このような要因の影響を踏まえた出力は再現されておらず、AI の出力はどこか画一的なものとなっている。塚野らによる「ニューラルネット言語モデルによる Twitter 上の発言からの 5 因子モデルに基づく性格分析」[1]のように、人間の持つ性格や感情を扱った研究はあるが、そこでも人の性格や感情を推定することに留まっており、それらを AI の出力に反映するまでには至っていない。

これに対し、本研究では、国籍によって受け止め方に差が生まれるような国際的なニューストピックスに対して生じる意思や感情の違いを出力に反映できる AI の構築を目的としている。ここでは、この違いを国民感情と呼ぶこととする。これを実現する事で、既存手法では成し得なかった国民感情を踏まえたような意見を出力する AI を構築できることが期待できる。

そのための一歩として、ある国際的なニューストピックスが、ある国籍の人にどのように受け止められるかを定義付けることにより、既存のニューラルネットワークモデルでも国民感情を学習できると仮説立て検証を行った。

本稿では、ニューラルネットワークを用いて国際的なニューストピックスとそれに向けられた意見の関係性を学習した結果、入力した国際的なニューストピックスに対して各国の立場から「良いニュース」、「悪いニュース」、「どちらでもないニュース」と意見表明するシステムを構築した事を報告する。

2. 提案システムの構築

本論文では、入力となる国際的なニューストピックスに対して、アメリカ、ロシア、中国それぞれの立場から、国民感情に基づいて「良いニュース」、「悪いニュース」、「どちらでもないニュース」と出力するニューラルネットワークを用いたシステムを提案する。そのために収集した学習データは、表 1 に示すように、国籍によって受け止め方に差が生まれるような国際的なニューストピックスを 100 件まとめたものである。また、教師データは、各国民が国民感情に基づいて表明した意見を One-hot 表現により、「良いニュース」、「悪いニュース」、「どちらでもないニュース」のいずれかを指すように示したものである。これをアメリカ、ロシア、中国それぞれの立場から 3 カ国分用意した。ここでのニューストピックスに対する「良い」、「悪い」、「どちらでもない」といった評価は、国際情勢を専門に研究している研究者からの意見を参考に筆者が定めたものである。

提案システムでは、収集した学習データを黒橋らが開発した形態素解析エンジンである Juman++[2]を用いて単語に分割した後に、word2vec[3]によって各単語を 50 次元のベクトルに変換することで、ニューラルネットワークへの入力とした。word2vec は、株式会社白ヤギコーポレーションが提供している学習済みのモデル[4]を利用した。

ニューラルネットワークの構成と訓練方法は表 2 に示す。入力層の 850 という数値は、学習データのうち、単語数が

表 1 学習データセットの例

学習データ	教師データ		
	アメリカ	ロシア	中国
中国が南シナ海にミサイル配備	悪い	どちらでも	良い
アメリカは環太平洋演習から中国を排除	悪い	悪い	悪い
アメリカは南シナ海に艦船を派遣	良い	どちらでも	悪い
アメリカはイラン核合意を離脱	悪い	良い	良い
ロシアが新型の大型魚雷を開発	悪い	良い	悪い

^{†1} 産業能率大学情報マネジメント学部
School of Information-Oriented Management, SANNO University

最も多いニューストピックスをベクトル化した時のニューストピックス全体の要素数である。ニューラルネットワークへの入力時、入力層のユニットが余る場合には、入力層の余分なユニットに0を入力した。構築したニューラルネットワークは、入力された国際的なニューストピックスを表すベクトルを中間層の64のユニットへ次元圧縮した後、softmax関数によって出力層の3つのユニットに格納する。この3つのユニットに格納された数値は、それぞれ「良いニュースである確率」、「悪いニュースである確率」、「どちらでもないニュースである確率」を示す。ニューラルネットワークは、このタスクを複数回繰り返す中で、教師データの示す出力層に格納されるべき正解となる値との誤差を最小にすることを目指し学習する。この誤差がより小さいほど、入力された国際的なニューストピックスと、教師データの示す各国の国民感情に基づいた意見との関係性を正確に学習したこととなる。

本研究では、3カ国の教師データそれぞれでニューラルネットワークの学習を行った。その結果、それぞれのニューラルネットワークは、図1~3に学習エポック数に対する誤差の推移を示す通り2000エポックの学習で誤差を0.05以下に収めた。このことから、国際的なニューストピックスとそれに向けられた国民感情を踏まえた意見の関係を学習したニューラルネットワークを3カ国分構築することができたと期待できる。次章では、構築したニューラルネットワークの精度を検証する。

表2 ニューラルネットワークの構成

入力層	850
中間層	64
出力層	3
活性化関数	ReLU・softmax
学習率	0.1
目的関数	crossentropy
最適化手法	SGD
学習エポック数	2000

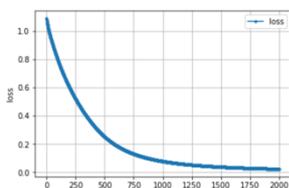


図1 アメリカ人の国民感情学習時の誤差推移

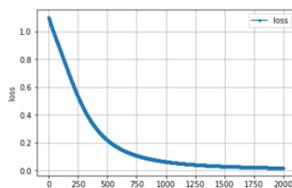


図2 ロシア人の国民感情学習時の誤差推移

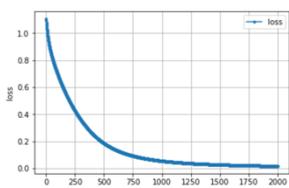


図3 中国人の国民感情学習時の誤差推移

3. 検証実験

3.1 テストデータの収集

検証実験では、構築した3カ国分のニューラルネットワークの精度をそれぞれ評価するため、「アメリカにとって良いとされるニューストピックス」、「ロシアにとって良いとされるニューストピックス」、「中国にとって良いとされるニューストピックス」をそれぞれ5トピックスずつ収集し、計3群のテストデータを用意した。これらのデータは、Yahooニュースの国際カテゴリから収集し、「どの国がどの国に対して何をした」という形に正規化した。また、学習データに含まれるニューストピックスとの重複は排除した。

3.2 検証実験

検証実験では、構築した3カ国のシステムの出力精度を3.1で記した3群のテストデータを用いて確認した。実験結果は、表3~5に各国民の国民感情を踏まえた出力をするように学習した場合毎に示す。表の左の列は、3.1で記したテストデータを示し、右の列は、各ニューラルネットワークの出力層に格納された各ユニットの値を小数点第5位で四捨五入した後示している。ここで期待するニューラルネットワークの出力は、学習した国民感情を持つ国民の国籍と、与えられたトピックスを良いとする国民の国籍が一致した場合に、出力層の「良い」を示すユニットが他のユニットより大きな数値を持つことである。また、学習した国民感情を持つ国民の国籍と、与えられたトピックスを良いとする国民の国籍が異なる場合には、出力層の「良い」以外を示すユニットが大きな数値を持つことを期待する。一方で、No.3のテストデータのように複数の国にとって良いとされるトピックスも存在しており、このようなテストデータに対して期待する出力は上記の限りではない。

これらを踏まえて結果を見ると、3カ国いずれの場合も高い精度で期待通りの出力を得ていることが分かる。しかしながら、課題として、表3のアメリカ人の国民感情を踏まえた出力をするように学習した場合は、No.5のテストデータで「良い」に格納された数値が0.2867と期待する出力を得られていないことが挙げられる。また、表4のロシア人の国民感情を踏まえた出力をするように学習した場合は、No.10のテストデータで「良い」に格納された数値が0.0461と期待する出力を得ることができていない。これら2つの課題に対して、国際的なニューストピックスに対する国民感情に基づいた意見表明は、ある国際的なニューストピックスがどこの国のトピックスであるかに大きく影響を受けることに対し、提案システムでは、各ニューラルネットワークの学習時に、学習データの示す国際的なニューストピックスがどこの国のトピックスであることを強く指示できていないことが原因であると考えた。本論文では、これを元に追加実験を行った。

表3 アメリカ人の国民感情を踏まえた出力をするよう学習したニューラルネットワークの出力結果

No	アメリカ国民に良いとされるトピックス	構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
1	アメリカは高関税による悪影響を懸念	0.9815	0.0014	0.0170
2	アメリカは北朝鮮の完全非核化まで制裁継続	0.9530	0.0278	0.0192
3	北朝鮮がミサイル実験場の解体に着手	0.8798	0.0465	0.0737
4	アメリカは対中関税一部製品除外へ	0.8709	0.0614	0.0677
5	アメリカは北朝鮮からの親書を公開	0.2867	0.2231	0.4902
ロシア国民にとって良いとされるトピックス				
		構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
6	ロシアがアメリカの北朝鮮制裁を阻止	0.0023	0.0063	0.9914
7	ロシアが新世代兵器を公開	0.0011	0.0004	0.9984
8	ロシアはアメリカ防衛国に報復措置を計画	0.1860	0.0495	0.7645
9	ロシアは北朝鮮との協力を中断	0.2044	0.1469	0.6487
10	アメリカはロシアに対する政策を見直し	0.9341	0.0634	0.0024
中国にとって良いとされるトピックス				
		構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
11	中国はアメリカの追加関税に報復措置取る	0.0005	0.2269	0.7726
12	中国と北朝鮮は在韓米軍の撤退を促す方針	0.0064	0.0321	0.9615
13	中国はアメリカの関税で自国の利益を保護	0.0145	0.4259	0.5597
14	中国はアメリカの輸入関税でWTOに不服	0.0306	0.8642	0.1052
15	アメリカは中国企業への制裁解除を発表	0.6901	0.3024	0.0076

表4 ロシア人の国民感情を踏まえた出力をするよう学習したニューラルネットワークの出力結果

No	アメリカ国民に良いとされるトピックス	構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
1	アメリカは高関税による悪影響を懸念	0.1551	0.1625	0.6823
2	アメリカは北朝鮮の完全非核化まで制裁継続	0.1558	0.6272	0.2171
3	北朝鮮がミサイル実験場の解体に着手	0.8720	0.0169	0.1111
4	アメリカは対中関税一部製品除外へ	0.9023	0.0182	0.0795
5	アメリカは北朝鮮からの親書を公開	0.0945	0.6945	0.2111
ロシア国民にとって良いとされるトピックス				
		構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
6	ロシアがアメリカの北朝鮮制裁を阻止	0.9975	0.0024	0.0001
7	ロシアが新世代兵器を公開	0.9823	0.0108	0.0070
8	ロシアはアメリカ防衛国に報復措置を計画	0.9646	0.0093	0.0261
9	ロシアは北朝鮮との協力を中断	0.8034	0.1341	0.0625
10	アメリカはロシアに対する政策を見直し	0.0461	0.8931	0.0608
中国にとって良いとされるトピックス				
		構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
11	中国はアメリカの追加関税に報復措置取る	0.2616	0.6907	0.0478
12	中国と北朝鮮は在韓米軍の撤退を促す方針	0.1610	0.8302	0.0088
13	中国はアメリカの関税で自国の利益を保護	0.2871	0.7036	0.0094
14	中国はアメリカの輸入関税でWTOに不服	0.1662	0.8075	0.0263
15	アメリカは中国企業への制裁解除を発表	0.6553	0.0636	0.2811

表5 中国人の国民感情を踏まえた出力をするよう学習したニューラルネットワークの出力結果

No	アメリカ国民に良いとされるトピックス	構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
1	アメリカは高関税による悪影響を懸念	0.0293	0.1078	0.8630
2	アメリカは北朝鮮の完全非核化まで制裁継続	0.4012	0.4068	0.1920
3	北朝鮮がミサイル実験場の解体に着手	0.9910	0.0006	0.0084
4	アメリカは対中関税一部製品除外へ	0.1174	0.6496	0.2330
5	アメリカは北朝鮮からの親書を公開	0.0928	0.0787	0.8285
ロシア国民にとって良いとされるトピックス				
		構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
6	ロシアがアメリカの北朝鮮制裁を阻止	0.9792	0.0158	0.0050
7	ロシアが新世代兵器を公開	0.2828	0.0303	0.6869
8	ロシアはアメリカ防衛国に報復措置を計画	0.1431	0.0382	0.8188
9	ロシアは北朝鮮との協力を中断	0.2403	0.4642	0.2956
10	アメリカはロシアに対する政策を見直し	0.0925	0.1086	0.7989
中国にとって良いとされるトピックス				
		構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
11	中国はアメリカの追加関税に報復措置取る	0.9995	0.0003	0.0002
12	中国と北朝鮮は在韓米軍の撤退を促す方針	0.9981	0.0004	0.0015
13	中国はアメリカの関税で自国の利益を保護	0.9879	0.0049	0.0072
14	中国はアメリカの輸入関税でWTOに不服	0.9310	0.0681	0.0009
15	アメリカは中国企業への制裁解除を発表	0.8482	0.0124	0.1394

表6 追加した学習データ

データセットに含まれる学習データ	データセットに含まれる教師データ		
	アメリカ	ロシア	中国
アメリカが中国の輸送機に軍事レーザー照射	良い	どちら	悪い
データセットに追加した学習データ	データセットに使用した教師データ		
	アメリカ	ロシア	中国
中国がアメリカの輸送機に軍事レーザー照射	悪い	どちら	良い

3.3 追加実験

検証実験で得られた知見を元に、追加実験では、学習データに含まれる国際的なニューストピックスが、どこの国のトピックスであるかをより強くニューラルネットワークに学習させるため、追加の学習データを与えた。追加した学習データは、学習データ内に含まれる特定の国における出来事を指すニューストピックスを他国が行なったものとして作り変えた偽ニュースである。追加した学習データの例は表6に示す。これにより、ニューラルネットワークの出力内容に国際的なニューストピックスがどこの国のトピックスであるかを大きく反映できると考えた。

追加実験の結果は、表7-9に3.2の結果と同様に示す。検証実験で課題に挙げたNo.5とNo.10の2つのテストデータを見ると、3.2で一つ目の課題として挙げたアメリカ人の国民感情を踏まえた出力をするように学習した場合のNo.5のテストデータ(表7)では、学習データを追加する以前に比べ、「良い」を示すユニットに格納された数値が極僅かに向上している。また、二つ目の課題として挙げたロシア人の国民感情を踏まえた出力をするように学習した場合のNo.10のテストデータ(表8)では、「良い」を示すユニットに格納された数値に0.4ポイントもの向上が確認できた。これに加えて、学習データを追加したことによって、その他の結果が悪影響を受けた様子は見られなかった。

以上のことから、追加実験で試行した手法は、3.2で挙げた課題に対して有効であると考えられる。しかしながら、依然として全てのテストデータで期待通りの出力を得るまでには至っていない。これを解決するためには、学習データ量を増加させることで、汎化性能の向上を目指すことが可能であると考えられる。

4. おわりに

本論文では、国際的なニューストピックスへの国民感情を踏まえた意見表明ができるAIの構築を目的に、ニューラルネットワークの構築と学習データの収集を行った。その結果、入力したニューストピックスをアメリカ、ロシア、中国の各国民の立場から、「良いニュース」、「悪いニュース」、「どちらでもないニュース」と意見表明するAIを構築することができた。このことから、AIの出力に国民感情を反映するためのアプローチとして、本稿で提案したニューラ

表7 アメリカ人の国民感情を踏まえた出力をするよう学習したニューラルネットワークの出力結果(改良後)

No	アメリカ国民に良いとされるトピックス	構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
1	アメリカは高関税による悪影響を懸念	0.9980	0.0001	0.0019
2	アメリカは北朝鮮の完全非核化まで制裁継続	0.9954	0.0017	0.0029
3	北朝鮮がミサイル実験場の解体に着手	0.9379	0.0128	0.0492
4	アメリカは対中間税一部製品除外へ	0.9666	0.0278	0.0056
5	アメリカは北朝鮮からの親書を公開	0.3187	0.2773	0.4040
ロシア国民にとって良いとされるトピックス				
	ロシア国民にとって良いとされるトピックス	構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
6	ロシアがアメリカの北朝鮮制裁を阻止	0.0050	0.0073	0.9877
7	ロシアが新世代兵器を公開	0.0008	0.0001	0.9992
8	ロシアはアメリカ防衛国に報復措置を計画	0.0021	0.0032	0.9948
9	ロシアは北朝鮮との協力を中断	0.0007	0.0049	0.9944
10	アメリカはロシアに対する政策を見直し	0.9207	0.0732	0.0061
中国にとって良いとされるトピックス				
	中国にとって良いとされるトピックス	構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
11	中国はアメリカの追加関税に報復措置取る	0.0012	0.0816	0.9172
12	中国と北朝鮮は在韓米軍の撤退を促す方針	0.0021	0.0139	0.9841
13	中国はアメリカの関税で自国の利益を保護	0.0092	0.2446	0.7462
14	中国はアメリカの輸入関税でWTOに不服	0.0478	0.5569	0.3954
15	アメリカは中国企業への制裁解除を発表	0.5005	0.4868	0.0127

表8 ロシア人の国民感情を踏まえた出力をするよう学習したニューラルネットワークの出力結果(改良後)

No	アメリカ国民に良いとされるトピックス	構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
1	アメリカは高関税による悪影響を懸念	0.0100	0.0096	0.9804
2	アメリカは北朝鮮の完全非核化まで制裁継続	0.0316	0.9026	0.0658
3	北朝鮮がミサイル実験場の解体に着手	0.9451	0.0125	0.0424
4	アメリカは対中間税一部製品除外へ	0.7540	0.0034	0.2426
5	アメリカは北朝鮮からの親書を公開	0.0216	0.7080	0.2703
ロシア国民にとって良いとされるトピックス				
	ロシア国民にとって良いとされるトピックス	構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
6	ロシアがアメリカの北朝鮮制裁を阻止	0.9997	0.0002	0.0000
7	ロシアが新世代兵器を公開	0.9995	0.0002	0.0001
8	ロシアはアメリカ防衛国に報復措置を計画	0.9997	0.0000	0.0001
9	ロシアは北朝鮮との協力を中断	0.9991	0.0008	0.0000
10	アメリカはロシアに対する政策を見直し	0.4195	0.5621	0.0182
中国にとって良いとされるトピックス				
	中国にとって良いとされるトピックス	構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
11	中国はアメリカの追加関税に報復措置取る	0.1734	0.8094	0.0172
12	中国と北朝鮮は在韓米軍の撤退を促す方針	0.0275	0.9710	0.0014
13	中国はアメリカの関税で自国の利益を保護	0.3790	0.5867	0.0343
14	中国はアメリカの輸入関税でWTOに不服	0.1731	0.7861	0.0407
15	アメリカは中国企業への制裁解除を発表	0.7377	0.0538	0.2085

表9 中国人の国民感情を踏まえた出力をするよう学習したニューラルネットワークの出力結果(改良後)

No	アメリカ国民に良いとされるトピックス	構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
1	アメリカは高関税による悪影響を懸念	0.0088	0.0163	0.9748
2	アメリカは北朝鮮の完全非核化まで制裁継続	0.2022	0.1686	0.6293
3	北朝鮮がミサイル実験場の解体に着手	0.9456	0.0005	0.0539
4	アメリカは対中間税一部製品除外へ	0.0929	0.4678	0.4393
5	アメリカは北朝鮮からの親書を公開	0.0200	0.0211	0.9589
ロシア国民にとって良いとされるトピックス				
	ロシア国民にとって良いとされるトピックス	構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
6	ロシアがアメリカの北朝鮮制裁を阻止	0.9419	0.0491	0.0090
7	ロシアが新世代兵器を公開	0.0697	0.1062	0.8241
8	ロシアはアメリカ防衛国に報復措置を計画	0.1031	0.0264	0.8705
9	ロシアは北朝鮮との協力を中断	0.2401	0.7306	0.0293
10	アメリカはロシアに対する政策を見直し	0.0052	0.0191	0.9757
中国にとって良いとされるトピックス				
	中国にとって良いとされるトピックス	構築したニューラルネットワークの出力		
		良い	どちらでもない	悪い
11	中国はアメリカの追加関税に報復措置取る	0.9999	0.0001	0.0000
12	中国と北朝鮮は在韓米軍の撤退を促す方針	1.0000	0.0000	0.0000
13	中国はアメリカの関税で自国の利益を保護	0.9998	0.0002	0.0000
14	中国はアメリカの輸入関税でWTOに不服	0.9667	0.0332	0.0001
15	アメリカは中国企業への制裁解除を発表	0.9550	0.0011	0.0440

ルネットワークによるアプローチが有効であることが分かった。

しかしながら、本稿で提案したシステムでは、学習データ数の不足と的確な教示方法を確立できていない事により、未知のテストデータに対して、国民感情に基づいて感化される意思や感情が異なるという人の意見表明プロセスを再現し切れていないという課題を残している。また、提案システムでは、出力できる意見は3つに限定されている。1章で示したような「人がある国際的なニューストピックスに対して表明する意見」は、本稿で提案したシステムでは再現できていない多様な意見が大半である。ニューラルネットワークへ与える学習データと教師データを多様化させることで汎化性能の向上と出力することのできる意見の幅を広げることができるが、これを実現する学習データの収集方法と的確なニューラルネットワークの学習方法が不明であるため、これらの解決を今後の展望とする。

5. 参考文献

- [1] 塚野俊, 柴田千尋, 政倉祐子, 田胡和哉: “ニューラルネットワークモデルによる Twitter 上の発言からの 5 因子モデルに基づく性格分析”, 情報処理学会第 78 回全国大会 (2016)
- [2] 黒橋 禎夫: “日本語形態素解析システム JUMAN++ version1.0 マニュアル” (2016), <http://lotus.kuee.kyoto-u.ac.jp/nl-resource/jumanpp/jumanpp-manual-1.0p2.pdf>
- [3] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J.: “Distributed Representations of Words and Phrases and their Compositionality”, Proc. 26th NIPS, pp. 3111–3119 (2013).
- [4] 株式会社白ヤギコーポレーション: “日本語 word2vec モデル” (2017), <http://public.shirovagi.s3.amazonaws.com/latest-japanese-word2vec-gensim-model.zip>
<https://github.com/shirovagicorp/japanese-word2vec-model-builder>