Homophily に基づくサイレントマジョリティの意見推定

向井 穂乃花 [1]　磯沼 大 [1]　森 純一郎 [1,2]　坂田 一郎 [1]
[1] 東京大学　[2] 理研 AIP
mukai-honoka377@g.ecc.u-tokyo.ac.jp　isonuma@ipr-ctr.t.u-tokyo.ac.jp
mori@mi.u-tokyo.ac.jp　isakata@ipr-ctr.t.u-tokyo.ac.jp

概要

本研究はソーシャルメディア上のサイレントユーザの意見を推定するための意見生成モデルを提案する。提案モデルでは、同じ意見を持つユーザはコミュニティを形成しやすいという “Homophily” に基づき、ユーザの意見をネットワーク埋め込みとトピックでモデル化する。これにより、サイレントユーザのネットワーク埋め込みをモデルに入力することで、様々なトピックに関する意見の推定を試みる。米国大統領選挙期間中のツイート文を対象に推定精度を検証した結果、既存の文生成モデルと同程度の性能が得られ、ネットワーク埋め込みがユーザの意見のモデル化に十分な情報を持つことが示唆された。更に潜在空間上のユーザの位置によって、ユーザの意見が変化していく様子が確認された。

1　はじめに

近年、ソーシャルメディア上の人々の意見は、個人の情報収集源として影響力を増しており、マスメディアを代替する情報源として活用されることも多い。一方、ソーシャルメディアは社会の分断に加担する負の側面もあり、自分と同様の価値観を持つ意見ばかりに囲まれるフィルターバブル現象などを通じて、ユーザを時に過激な言動に走らせる [1, 2, 3]。しかし、実際には過激なユーザは少数で、多数は穏健な考えを持つが意見表明を積極的に行わないサイレントマジョリティであり、Twitter 上では約 40% のユーザがサイレントユーザだという報告もある [4, 5]。サイレントマジョリティの意見推定は、ソーシャルメディアを情報収集源として活用すると共に、人々の視野を広げ過激な行動を防ぐための重要なタスクである。

サイレントマジョリティの意見推定の文脈において、既存研究 [6] では同じ意見や態度を持つユーザはコミュニティを形成しやすいという “Homophily”[7] という性質を活用し、サイレントユーザの感情極性推定を試みている。しかし同じ感情極性を示していても、その根拠が異なる場合は多い。例えば米国で反共和党を掲げる人々も、その理由は白人至上主義的な考えへの反発や、COVID-19 への対応策への批判など多岐にわたる。こうした思考プロセスを含めた意見を把握するには、挑戦的であるものの極性分類に留まらず文生成による意見推定が望ましい。

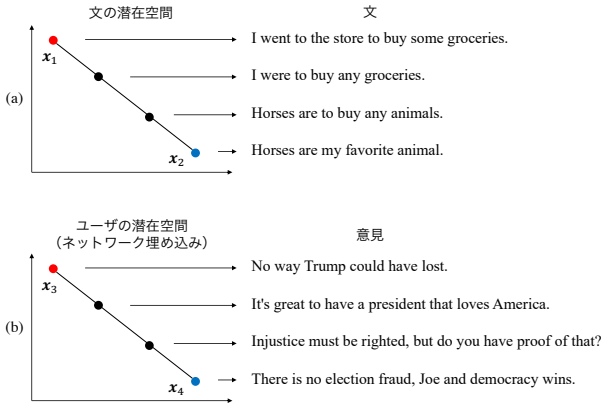

図 1: Bowman らの文生成モデル (a) と
提案する意見生成モデル (b) の対比

そこで、本研究は Bowman らによる文生成モデル [8] を応用することで、サイレントユーザの意見推定を試みる。図 1 にその概念図を示す。Bowman らは、文の潜在表現がガウス分布に従うと仮定し、Variational Autoencoder (VAE[9, 10]) の枠組みで文とその潜在表現間の写像を学習している。これにより、例えば図 1（a）のような 2 次元の潜在空間が与えられた時、$\boldsymbol{x}_1, \boldsymbol{x}_2$ 間の等間隔上に位置する潜在表現からは、$\boldsymbol{x}_1, \boldsymbol{x}_2$ から復号された文に対し意味が少しずつ変化した文が復号されることを明らかにした。一方、本研究はソーシャルメディア上のユーザの潜在表現をネットワーク埋め込みで獲得し、ユーザの潜在空間上における意見分布をモデル化する。Homophily を仮定すると、潜在空間上で近接す

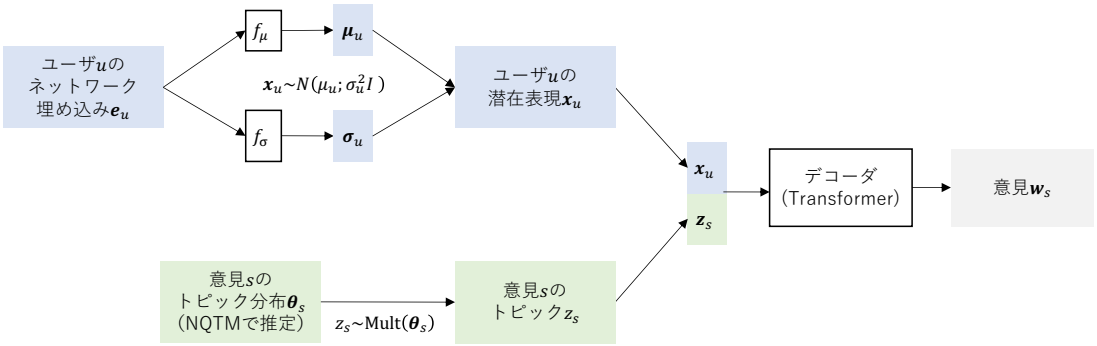

図 2: 提案モデルの概要図

るユーザ（ネットワーク上で近接するユーザ）は類似する意見を持つため、ユーザの潜在表現と意見間の写像を学習することで、サイレントユーザの意見を推定できると考えた。例えば図 1（b）における $\boldsymbol{x}_3, \boldsymbol{x}_4$ のように顕在化しているユーザの意見を学習することで、その中間に位置するサイレントユーザの意見推定を図る。

評価実験では、米国大統領選挙期間中のツイート文をユーザの意見とみなし、ユーザのネットワーク埋め込み表現をもとにツイート文を推定する実験を行った。評価データセット中のツイート文について Bowman らの文生成モデルと同程度の尤度が得られることを確認し、サイレントユーザの意見推定に提案法が有効であることが示唆された。

2 事前準備

まず、本研究の基礎となる Bowman らによる文生成モデル（VAE–LM）について解説する。Bowman らは文 $\boldsymbol{w}_s$ の生成過程を以下でモデル化している。

1. 文の潜在表現 $\boldsymbol{x}_s \in \mathbb{R}^d$ をサンプル:

$$\boldsymbol{x}_s \sim \mathcal{N}(\boldsymbol{x}_s | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \tag{1}$$

2. $\boldsymbol{x}_s$ を復号し、文 $\boldsymbol{w}_s$ をサンプル：

$$\boldsymbol{w}_s | \boldsymbol{x}_s \sim p(\boldsymbol{w}_s | \boldsymbol{x}_s) = \mathrm{Decoder}(\boldsymbol{x}_s) \tag{2}$$

ただし、$\boldsymbol{\mu}_0 = \boldsymbol{0}, \boldsymbol{\Sigma}_0 = \boldsymbol{I}$ はガウス分布のパラメータである。この生成モデルを VAE の枠組みで学習することによって、図 1（a）に示すように、2 つの文の潜在表現の中間点から、2 文と意味が似た文を復号できることを明らかにした。

これを発展させ、Tang らは文の潜在表現をトピック z_s と文構造などその他の特徴 $\boldsymbol{x}_s$ に分解できると仮定し、文の生成過程を以下でモデル化した（TATGM; Topic Augmented Text Generation Model[11]）。

1. 文の潜在表現 $\boldsymbol{x}_s \in \mathbb{R}^d$ をサンプル:

$$\boldsymbol{x}_s \sim \mathcal{N}(\boldsymbol{x}_s | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \tag{3}$$

2. 文のトピック $z_s \in \{1, \ldots, K\}$ をサンプル:

$$z_s \sim \mathrm{Cat}(\boldsymbol{\theta}_s) \tag{4}$$

3. $\boldsymbol{x}_s, z_s$ を復号し、文 $\boldsymbol{w}_s$ をサンプル：

$$\boldsymbol{w}_s | \boldsymbol{x}_s, z_s \sim p(\boldsymbol{w}_s | \boldsymbol{x}_s, z_s) = \mathrm{Decoder}([\boldsymbol{x}_s; z_s]) \tag{5}$$

ただし、$\boldsymbol{\theta}_s \in \Delta^{K-1}$ は文 s のトピック分布で、Neural Topic Model[12, 13] により推定する。また、$[\boldsymbol{x}_s; z_s]$ は $\boldsymbol{x}_s$ と z_s の one-hot 表現 $\boldsymbol{z}_s$ を結合したベクトルである。トピックを陽にモデル化することで、$\boldsymbol{x}_s$ を固定して z_s を変化させると、文構造などは変化させずに、トピックのみを変化させることを可能にした。

一方、本研究は TATGM をベースに、文の潜在表現 $\boldsymbol{x}_s$ をユーザの潜在表現（ネットワーク埋め込み表現）$\boldsymbol{x}_u$ に置き換える。これにより、ソーシャルネットワーク上の位置によって特徴づけられたあるユーザが、任意のトピックを与えられたときに、どのような意見を発するかについてモデル化する。

3 提案モデル

提案モデル概要を図 2 に示す。本研究では、ユーザ u の意見 $\boldsymbol{w}_s$ を以下の生成過程でモデル化する。

・各ユーザ $u \in \{1, \ldots, U\}$ について:

1. ユーザの潜在表現 $\boldsymbol{x}_u \in \mathbb{R}^d$ をサンプル:

$$\boldsymbol{x}_u \sim \mathcal{N}(\boldsymbol{x}_u | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \tag{6}$$

・ユーザ u の各意見 $s \in \{1, \ldots, S_u\}$ について:

2. 意見のトピック $z_s \in \{1, \ldots, K\}$ をサンプル:

$$z_s \sim \mathrm{Cat}(\boldsymbol{\theta}_s) \tag{7}$$

3. $\boldsymbol{x}_u, z_s$ を復号し、意見 $\boldsymbol{w}_s$ をサンプル：

$$\boldsymbol{w}_s | \boldsymbol{x}_u, z_s \sim p(\boldsymbol{w}_s | \boldsymbol{x}_u, z_s) = \mathrm{Decoder}([\boldsymbol{x}_u; z_s]) \tag{8}$$

このとき、ユーザ u の意見の尤度と、その対数の変分下限 $\mathcal{L}_u$ はそれぞれ以下で表される。

$$p(\boldsymbol{W}_{1:S_u}) = \int p(\boldsymbol{x}_u) \prod_{s=1}^{S_u} \left\{ \sum_{z_s} p(\boldsymbol{w}_s|\boldsymbol{x}_u, z_s) p(z_s) \right\} d\boldsymbol{x}_u \quad (9)$$

$$\begin{aligned} \mathcal{L}_u \geq & \mathbb{E}_{q(\boldsymbol{x}_u|\boldsymbol{e}_u)} \left\{ \sum_{s=1}^{S_u} \log p(\boldsymbol{w}_s|\boldsymbol{x}_u, \boldsymbol{\theta}_s) \right\} \\ & - \mathrm{D}_{\mathrm{KL}} \left[q(\boldsymbol{x}_u|\boldsymbol{e}_u) || p(\boldsymbol{x}_u) \right] \end{aligned} \quad (10)$$

ただし、$q(\boldsymbol{x}_u|\boldsymbol{e}_u) = N(\boldsymbol{x}_u|f_\mu(\boldsymbol{e}_u), f_\sigma(\boldsymbol{e}_u))$ はユーザ u の潜在表現の変分事後分布で、ユーザのネットワーク埋め込み表現 $\boldsymbol{e}_u \in \mathbb{R}^h$ を MLP f_μ, f_σ により変換することで推定する。また、$\boldsymbol{\theta}_s \in \Delta^{K-1}$ は文 s のトピック分布であり、本研究では Neural Topic Model の一種である NQTM[14] により獲得する。以下、それぞれを用いた理由とその方法について概説する。

3.1 ユーザのネットワーク埋め込み

本研究はサイレントユーザの意見推定を試みることから、ユーザの意見（テキスト）に頼らずにユーザの潜在表現を得る必要がある。そこで、Homophily の概念に着想を得て、ユーザのネットワーク埋め込み表現を用いることで潜在表現を獲得する。

本研究ではユーザをノード、ユーザ同士のコミュニケーション（リツイート/リプライ）数をエッジとした重み付き無向グラフを対象に、ネットワーク埋め込み手法として LINE[15] を用いた。LINE では同一コミュニティに属するユーザが近接するように埋め込み表現を学習する。従って、ネットワーク埋め込みが近接するユーザ群からは、類似する意見が生成され、Homophily をモデル化できると期待される。

3.2 NQTM による意見のトピック推定

TATGM ではトピック分布の推定に Neural Topic Model[12] を用いているが、本研究が対象とするソーシャルメディア上の投稿など、短文のテキストにはトピック推定がうまく機能しない。そこで、本研究では短文テキストを対象に開発された NQTM（Negative sampling and Quantization Topic Model[14]）を利用する。短文のテキストでは、トピック分布が一様分布に近くなるとともに、各トピックの単語分布が似た分布になりやすい。そこで NQTM では、トピック分布の量子化により鋭いピークを持つ分布に誘導すると共に、他トピックに頻出する単語を負例として用いるネガティブサンプリングによって、各トピックの単語分布の差が際立つよう学習する。

表 1: データセット中のツイート数とユーザ数

データセット	ツイート数	ユーザ数
学習	8,439,151	1,400,829
検証	1,852,497	708,325
評価	3,422,534	499,415

表 2: 評価データの意見推定精度（低いほうがよい）

モデル	負の対数尤度
VAE–LM (Bowman et al.; [8])	127.69
TATGM (Tang et al.; [11])	112.71
提案モデル	116.61

4 実験

本実験では、ソーシャルメディアとして Twitter を採用し、Twitter 上の投稿（ツイート文）を意見とみなすことで、提案モデルが各ユーザの意見をどの程度モデル化できるか、即ちサイレントユーザの意見をどの程度推定可能か検証する。

4.1 実験設定

本実験では、政治的意見が多く含まれると考えられる 2020 年アメリカ大統領選挙の前後期間に特定のクエリで抽出されたツイート文 [16] を対象に、意見生成の精度評価を行った。データセット中のツイート数/ユーザ数を表 1 に示す。ベースラインとして、Bowman らの VAE–LM[8] と、Tang らの TATGM[11] を採用した。トピック数は 10 で固定し、ベースライン・提案モデル共に Transformer[17] をデコーダとして用いた。

4.2 意見推定の精度比較

まず、提案モデルが各ユーザの意見をどの程度モデル化できるか検証するため、評価データセット中のツイート文の尤度を確認した。学習に利用していない評価データ中のユーザはいわばサイレントユーザとみなすことができるため、尤度はサイレントユーザの意見推定精度とみなせる。各モデルから算出されたツイート文の負の対数尤度を表 2 に示す。提案モデルは Bowman より高い推定精度を確認できた他、TATGM にも匹敵する性能を示した。VAE–LM や TATGM は推定対象の文 $\boldsymbol{w}_s$ を用いて、文の潜在表現の事後分布を推定している：$q(\boldsymbol{x}_s|\boldsymbol{w}_s) = N(\boldsymbol{x}_u|f_\mu(\boldsymbol{w}_s), f_\sigma(\boldsymbol{w}_s))$（ただし、$f_\mu, f_\sigma$ は Transformer エンコーダ）。一方、提案モデルは事後分布の推定に文 $\boldsymbol{w}_s$ を用いず、ネットワーク埋め込みを用いているため、ベースラインの方が文 $\boldsymbol{w}_s$ の

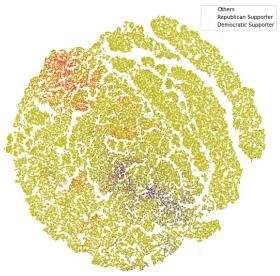

図 3: ユーザの潜在空間

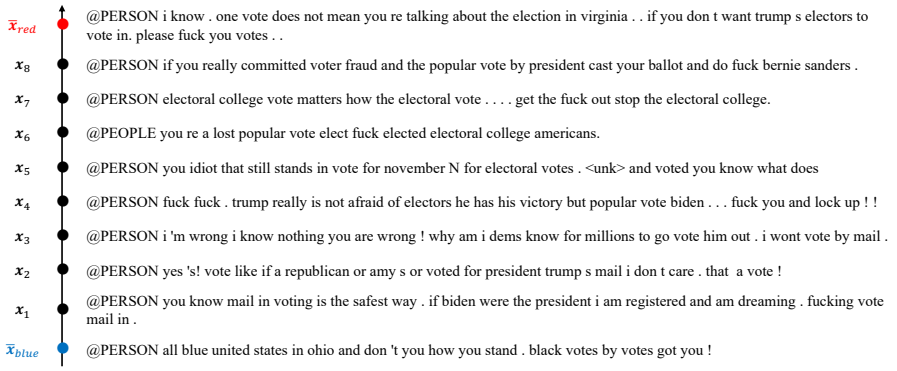

図 4: ユーザの潜在変数を移動した時のトピック 1 に関する生成文変化

表 3: 各トピックにおける頻出度上位 10 単語

トピック	頻度上位 10 単語
1	electoral vote votes win election nope fraud supreme democrats majority
2	deaths trump virus covid pandemicdeath cases coronavirus testing gaiters americans
3	blah fake trump anything news shit lies mouth lie liar
4	parole brandonbernard commute bernard sentence brandon paro please parol violates
5	bless birthday god president happy trump thank merry donald best
6	absentee ballot ballots mail counted votes electoral mailed vote voting
7	yawn womp ouch nailed tock yikes boom cuck tick yuck
8	hunter peopleschamp tNa reckoningreckoning elect reade questions president vice pN
9	giphy randomtrump court fraud supreme probe voter papadopoulos election capitol
10	antifa supporters blm terrorist trump supremacists protesters supremacist terrorists police

推定は容易であると考えられる。それにも関わらず提案モデルがベースラインと競合する性能を得られたことは、ネットワーク埋め込みがユーザの意見のモデル化に十分な情報を持つことを示唆している。

4.3 トピックの分析

表 3 に各トピックに最も頻出する上位 10 単語を示す。総じて、選挙戦を中心とした政治的なトピックに関する単語が取得できていることがわかる。特にトピック 1,3,5,6,8,9 は政局や選挙関連について激しい論戦が行われた話題である。トピック 1 は不正投票、トピック 2 は COVID-19、トピック 4 は死刑判決が議論を呼んだ Brandon Bernard 氏、トピック 10 は 2020 年に米国で盛んだった Black Lives Matter 運動について、それぞれ示していると考えられる。

4.4 ユーザの潜在空間の分析

評価データセット中のユーザの潜在表現 $\boldsymbol{x}_u$ について、t-SNE[18] による 2 次元空間上への射影を図 3 に示す。各点はユーザを表し、共和党/民主党支持に関連するハッシュタグ（付録 A 表 4）どちらかを使用したユーザをそれぞれ赤/青で表している。両者は互いに固まって位置していることが確認できる。

また潜在空間内の共和党支持ユーザ（赤）と民主党支持ユーザ（青）の重心ベクトル $\bar{\boldsymbol{x}}_{red}, \bar{\boldsymbol{x}}_{blue}$ の中間地点のユーザベクトルをモデルに入力した時の出力文の変化例を図 4 に示す。ここでは $z_s = 1$ を入力し、不正選挙に関するトピック 1 について、その意見の変化を観察した。米大統領選では、選挙不正に関して郵便投票と選挙人団入替えの是非が主な争点であった。$\bar{\boldsymbol{x}}_{blue}$ に近づくほど、郵便投票に対し肯定的な意見が生成されている一方（$\boldsymbol{x}_1$）、$\bar{\boldsymbol{x}}_{red}$ に近づくと選挙人団投票 (electoral vote) に対する否定的な意見が生成される（$\boldsymbol{x}_7$）傾向を確認できる。このように潜在空間上の位置によって、ユーザの意見が変化していく様子が確認された。

5 おわりに

本研究はユーザの意見をネットワーク埋め込みとトピックに基づきモデル化することによって、様々なトピックに関するサイレントユーザの意見を推定できるモデルを提案した。米国大統領選挙期間中の Twitter データで意見（ツイート文）の推定精度を検証した結果、既存の文生成モデルと同程度の性能が得られた他、ユーザの潜在変数が各党の支持ユーザに近づくと、各党の支持ユーザらしい意見が生成される様子が定性的にも確認された。以上のことから、ネットワーク埋め込みとトピックに基づくユーザの意見のモデル化の妥当性が確認された。

謝辞

本研究は、JST ACT-X JPMJAX1904 及び JST CREST JPMJCR21D1 の支援を受けたものである。

参考文献

[1] Eli Pariser. **The filter bubble: How the new personalized web is changing what we read and how we think**. Penguin, 2011.

[2] Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 us election: divided they blog. In **Proceedings of the 3rd International Workshop on Link discovery**, pp. 36–43, 2005.

[3] Elanor Colleoni, Alessandro Rozza, and Adam Arvidsson. Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data. **Journal of Communication**, Vol. 64, No. 2, pp. 317–332, 2014.

[4] Yoree Koh. Report: 44% of twitter accounts have never sent a tweet. **Wall Street Journal. News Corporation**, Vol. 11, , 2014.

[5] Wei Gong, Ee-Peng Lim, and Feida Zhu. Characterizing silent users in social media communities. In **Proceedings of the International AAAI Conference on Web and Social Media**, Vol. 9, 2015.

[6] Lei Wang, Jianwei Niu, Xuefeng Liu, and Kaili Mao. The silent majority speaks: Inferring silent users' opinions in online social networks. In **The World Wide Web Conference**, pp. 3321–3327, 2019.

[7] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. **Annual Review of Sociology**, Vol. 27, No. 1, pp. 415–444, 2001.

[8] Samuel Bowman, Luke Vilnis, Oriol Vinyals, Andrew Dai, Rafal Jozefowicz, and Samy Bengio. Generating sentences from a continuous space. In **Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning**, pp. 10–21, 2016.

[9] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In **Proceedings of the 2nd International Conference on Learning Representations**, 2014.

[10] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In **Proceedings of the 31st International Conference on Machine Learning**, pp. 1278–1286, 2014.

[11] Hongyin Tang, Miao Li, and Beihong Jin. A topic augmented text generation model: Joint learning of semantics and structural features. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing**, pp. 5090–5099, 2019.

[12] Yishu Miao, Edward Grefenstette, and Phil Blunsom. Discovering discrete latent topics with neural variational inference. In **Proceedings of the 34th International Conference on Machine Learning**, pp. 2410–2419, 2017.

[13] Akash Srivastava and Charles Sutton. Autoencoding variational inference for topic models. In **Proceedings of the 5th International Conference on Learning Representations**, 2017.

[14] Xiaobao Wu, Chunping Li, Yan Zhu, and Yishu Miao. Short text topic modeling with topic distribution quantization and negative sampling decoder. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing**, pp. 1772–1782, 2020.

[15] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In **Proceedings of the 24th International Conference on World Wide Web**, pp. 1067–1077, 2015.

[16] Emily Chen, Ashok Deb, and Emilio Ferrara. # election2020: the first public twitter dataset on the 2020 us presidential election. **Journal of Computational Social Science**, pp. 1–18, 2021.

[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In **Advances in Neural Information Processing Systems**, pp. 5998–6008, 2017.

[18] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. **Journal of Machine Learning Research**, Vol. 9, No. 11, 2008.

[19] Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P Xing. Toward controlled generation of text. In **Proceedings of the 34th International Conference on Machine Learning**, pp. 1587–1596, 2017.

[20] Chris J Maddison, Andriy Mnih, and Yee Whye Teh. The concrete distribution: A continuous relaxation of discrete random variables. In **Proceedings of the 5th International Conference on Learning Representations**, 2017.

[21] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. In **Proceedings of the 5th International Conference on Learning Representations**, 2017.

[22] Victor Prokhorov, Ehsan Shareghi, Yingzhen Li, Mohammad Taher Pilehvar, and Nigel Collier. On the importance of the kullback-leibler divergence term in variational autoencoders for text generation. In **Proceedings of the 3rd Workshop on Neural Generation and Translation**, pp. 118–127, 2019.

[23] Bohan Li, Junxian He, Graham Neubig, Taylor Berg-Kirkpatrick, and Yiming Yang. A surprisingly effective fix for deep latent variable modeling of text. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing**, pp. 3603–3614, 2019.

A 付録

A.1 モデル詳細

本研究では、生成意見の対数尤度の変分下限 $\mathcal{L}_u$(10) に、下記の節で定義する $\mathcal{L}_t$ 及び $\mathcal{L}_d$ を加えた目的関数 $\mathcal{L}$ を最大化するようにモデルを学習する。

$$\mathcal{L} = \mathcal{L}_u + \mathcal{L}_t + \lambda \mathcal{L}_d \tag{11}$$

A.1.1 Neural Topic Model 及び NQTM

本節では TATGM にて用いられている Neural Topic Model[12, 13] と、それを短文のテキストに適用できるよう改良された NQTM[14] について概説する。Neural Topic Model では一般的なトピックモデルと同様に以下の生成過程を仮定する。

・ユーザ u の各意見 $s \in \{1, \ldots, S_u\}$ について:

1. 意見 s のトピック分布 $\boldsymbol{\theta}_s \in \Delta^{K-1}$ をサンプル:

$$\boldsymbol{y}_s \sim \mathcal{N}(\boldsymbol{y}_s | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \tag{12}$$

$$\boldsymbol{\theta}_s = \mathrm{softmax}(\boldsymbol{y}_s) \tag{13}$$

・意見 s の各単語 $n \in \{1, \ldots, N_s\}$ について:

2. 単語のトピック $z_n \in \{1, \ldots, K\}$ をサンプル:

$$z_n \sim \mathrm{Cat}(\boldsymbol{\theta}_s) \tag{14}$$

3. 単語 $w_n \in \{1, \ldots, V\}$ をサンプル：

$$w_n \sim \mathrm{Cat}(\boldsymbol{\beta}_{z_n}) \tag{15}$$

このとき、トピックモデルにおけるユーザ u の対数尤度の変分下限は以下で与えられる。

$$\mathcal{L}_t = \sum_{s=1}^{S_u} \sum_{n=1}^{N_s} \log(\boldsymbol{\beta} \cdot \hat{\boldsymbol{\theta}})_{w_n} - \mathrm{D}_{\mathrm{KL}}\left[q(\boldsymbol{y}_s | \boldsymbol{w}_s) || p(\boldsymbol{y}_s)\right] \tag{16}$$

ただし、$q(\boldsymbol{y}_s | \boldsymbol{w}_s) = q(\boldsymbol{y}_s | f_\mu(\boldsymbol{w}_s), f_\Sigma(\boldsymbol{w}_s))$ は変分事後分布であり、文 $\boldsymbol{w}_s$ の bag-of-words 表現を MLP f_μ 及び f_σ で変換することで得られる。NQTM では、変分下限 $\mathcal{L}_t$ に加え、負例の尤度及び量子化前後のトピック分布の誤差項を $\mathcal{L}_t$ に加えている。

A.1.2 Discriminator

TATGM[11] と同様に、生成された文 $\hat{\boldsymbol{w}}_s$ が指定されたトピック z_s に関する内容を含むように誘導するため、本研究では提案モデルに discriminator[19] を加えた。Discriminator は、生成文のサンプルをトピックモデルに入力して得られるトピック分布 $q(z_s | \hat{\boldsymbol{w}}_s)$ が、元の文のトピック分布 $p(z_s)$ に近づくように、以下の目的関数を最大化する。

表 4: ハッシュタグの分類

支持政党	ハッシュタグ
共和党	#maga #trump2020 #maga2020 #obamagate #kag #kag2020 #obamagate #donaldtrump #votered #voteredtosaveamerica
民主党	#bidenharris2020 #votebidenharris2020 #votebluetosaveamerica #voteblue #bidenharris2020 #votebluetoendthisnightmare #trumphasnoplan #trumpvirus #trumpisanationaldisgrace #trumpislosing #trumpisacompletefailure #dumptrump2020

$$\mathcal{L}_d = \sum_{s=1}^{S_u} \mathbb{E}_{p(\hat{\boldsymbol{w}}_s, z_s)}[\log q(z_s | \hat{\boldsymbol{w}}_s)] \tag{17}$$

$$\approx \sum_{s=1}^{S_u} \sum_{k=1}^{K} \log q(z_s = k | \hat{\boldsymbol{w}}'_s) \tag{18}$$

ただし、$\hat{\boldsymbol{w}}'_s$ は Gumbel-softmax trick[20, 21] によって推定した生成文のサンプルである。

A.2 実験詳細

A.2.1 各党支持に関連するハッシュタグ

各党の支持ユーザを同定するために、本研究では表 4 に示すハッシュタグを用いた。民主党の方がハッシュタグが多いものの、共和党/民主党支持と判断されたユーザ数はそれぞれ 15,961/9,274 である。なお、*#trump* や*#biden* など、肯定的/否定的な文脈双方で頻用されるものについては使用していない。

A.2.2 実装詳細

図 2 の f_μ, f_σ は 4 つの全結合層とバッチ正規化から成り、以下の式のように $\boldsymbol{\mu}_u, \boldsymbol{\sigma}_u$ を得る。

$$\boldsymbol{c}_0 = \boldsymbol{W}_0 \boldsymbol{x} + \boldsymbol{b}_0 \tag{19}$$

$$\boldsymbol{c}_1 = f_b(\boldsymbol{W}_1 \boldsymbol{c}_0 + \boldsymbol{b}_1) \tag{20}$$

$$\boldsymbol{\mu} = \boldsymbol{W}_2 \boldsymbol{c} + \boldsymbol{b}_2 \tag{21}$$

$$\boldsymbol{\sigma}^2 = \mathrm{softplus}(\boldsymbol{W}_3 \boldsymbol{c} + \boldsymbol{b}_3) \tag{22}$$

ただし $(\boldsymbol{W}_0, \boldsymbol{b}_0)$ は 128 次元ベクトルに、$(\boldsymbol{W}_1, \boldsymbol{b}_1), (\boldsymbol{W}_2, \boldsymbol{b}_2), (\boldsymbol{W}_3, \boldsymbol{b}_3)$ は 64 次元ベクトルに変換する全結合層のパラメータを表し、f_b はバッチ正規化関数を表す。デコーダの Transformer の隠れ層は 256、ブロック数は 6 に設定した。また単語ベクトルの次元数は 256、学習時のバッチサイズは 64、学習率は 0.0001、勾配降下法は Adam($(\beta_1, \beta_2) = (0.9, 0.999)$, weight_decay= 0)、Discriminator の係数は $\lambda = 0.1$ を採用した。また、VAE の事後分布崩壊を防ぐために、KL-annealing 及び KL-threshold[22, 23] を適用した。