

# A Note on Tweet Analysis for Extracting and Visualizing Service Improvement Ideas

Natsuki HASHIMOTO<sup>\*</sup>, Hiroyoshi NAGAO<sup>\*\*</sup> and Marie KATSURAI<sup>\*\*</sup>

(Received October 14, 2022)

Service developers are constantly working to improve their services. Although opinions and feedback from users are important for improving their services, conducting an interview is time consuming. Thus, this paper proposes an approach for extracting ideas from service users' tweets. The proposed method consists of tweet collection, textual feature extraction, and visualization, which are easy to implement. In experiments targeting three famous chat services, we demonstrated the effectiveness of the proposed method compared with machine learning methods. We also provide a qualitative analysis of the extracted ideas via our visualization interface constructed to understand the tweet analysis results intuitively.

Key words : idea extraction, service improvement, tweet analysis, visualization

キーワード : アイデア抽出, サービス改善, ツイート分析, 可視化

## サービス改善アイデアの抽出と可視化に向けた ツイート分析手法の検討

橋本 夏希, 長尾 浩良, 桂井 麻里衣

### 1 はじめに

サービスの開発者らは、既存機能の改良や新たな価値の提供に向けて日々努力を続けている。ユーザに長く使ってもらえるようなサービスを考案するにはユーザ目線での発想が必要不可欠である。自社サービスの改善に際しては、実際に利用したユーザの意見やフィードバックが重要であるが、インタビューの実施は人的コストが大きい。そこで本研究では、ソーシャルネットワーキングサービスから自動で意見抽出することを提案する。最も有名なマイクロブログサービスである Twitter は、一日のアクティブユーザ数が 2 億 3780 万人<sup>1)</sup> (2022 年 7 月時点) であり、消費者が自分の希望や率直な意見をつ

ぶやくことが多い。これらのツイートから既存サービスを改善するためのヒントが得られると考えられる。

関連研究として、ツイートからのオピニオンマイニング<sup>2)</sup>が挙げられる。オピニオンマイニングとは、ある特定の事柄に関する検索結果の集合において、その事柄の属性(品質、特徴など)のリストを生成し、それぞれに対する意見(悪い、普通、良い)を集計することを指す<sup>3)</sup>。しかしながら、これらの研究はある事柄に対するユーザの感情分析に注目しており、サービス改善策のように具体的なアイデアを抽出する手法はほとんど研究されていない。近年では、ツイートから特定の事象に関する人々のニーズ情報(例: COVID-19 でのニューヨーク州の人々の心理的ニーズ<sup>4)</sup>)を抽出する研究が注目され

<sup>\*</sup> Department of Information and Computer Science, Graduate School of Science and Engineering, Doshisha University, Kyoto  
E-mail : natsuki.hashimoto@mm.doshisha.ac.jp

<sup>\*\*</sup> Department of Intelligent Information Engineering and Sciences, Faculty of Science and Engineering, Doshisha University, Kyoto  
E-mail : {nagao21,katsurai}@mm.doshisha.ac.jp

始めたが、企業のサービス改善のようなビジネスツールとしての提案は未だ検討が少ない。

本研究では、サービス改善アイデアの抽出・可視化に向けたツイート分析手法として、簡便なツイート収集および単語分析フレームワークを提案する。実存サービスを対象とした実験では、より複雑な機械学習手法との比較に基づき、提案手法の有効性を示す。さらに、分析結果をより直感的に理解するための可視化インタフェースを紹介する。

## 2 アイデア抽出のための単語分析および可視化

### 2.1 分析用ツイートの収集手法

まず、Twitter API<sup>a</sup>を用いて分析用の日本語ツイートを収集する。対象サービスの改善案を含むツイートを得るために、(i) 対象のサービス名の文字列（「LINE」「Weibo」など）と (ii) 「機能欲しい」という文字列の両方を含むツイートを収集する。他ユーザのリツイートなど、重複するツイートは除外する。

収集したツイートには対象サービス以外のサービスについて言及しているツイートも含まれる。例えば「LINE」と「機能欲しい」を含むツイートを収集した際、「LINEみたいなビデオ通話機能が欲しい」のように、LINEへの改善案ではないツイートも収集される。そのため、簡便なフィルタリングとして、「(サービス名) みたい」を含むツイートを収集対象から除外する。

### 2.2 ツイートの単語分析手法

ツイートを収集するだけでは、どのようなアイデアが含まれているかを把握することは難しい。そのため、類似した意見のツイートをまとめることで、異なるアイデアの内容を把握できるようにする。提案手法では、まず各ツイートのテキストをベクトルで表現する。近年の自然言語処理でよく用いられる文の埋め込み<sup>5)</sup>は、意味を抽象化して抽出したい場面には適するが、ここで我々が着目したいのはサービスに関する具体的なキーワードである。そのため、単語単位でのベクトル化として、term frequency-inverse document frequency (TF-IDF) によるツイート集約を採用する。具体的には、まず全ツイートに対して GiNZA<sup>6)</sup>を用いて形態素解析を行い、名詞、形容詞、動詞、副詞、固有名詞を抽出する。これらの品詞に該当する単語をもとに TF-IDF ベクトルを作成し、

ベクトル間のコサイン類似度を算出する。類似度が予め設定した閾値を超える場合、対応する2つのツイートを類似度が高いペアとみなして保存する。

### 2.3 アイデアクラスタの可視化

前節で抽出したツイートペアの効果的な可視化として、ノードが各ツイート、エッジがツイート間の類似関係を表すようなツイートネットワークを構築する。加えて、ツイート集合に含まれるアイデア内容を一瞥で理解できるようにするために、ワードクラウドを用いる。可視化インタフェースの構築には Dash<sup>b)</sup>を用いる。Dash は、Plotly.js と React.js の上に書かれた Python のフレームワークであり、インタラクティブなデータ可視化機能を持つ Web アプリを構築することができる。

## 3 実験と検討

有名なチャットサービスである LINE<sup>c)</sup>、Slack<sup>d)</sup>、Discord<sup>e)</sup>を対象にアイデア抽出実験を実施した。これらの競合する3つのサービスに関して意見を比較することはそれぞれの長所・短所の分析につながるため、全サービスの分析結果を単一のインタフェースで可視化する。

2021年1月1日から2021年12月31日までに投稿されたツイートのうち、サービス名と「機能欲しい」という文字列を含むツイートは、LINEが1,243件、Slackが106件、Discordが216件、計1,565件であった。提案手法によるフィルタリング適用後には、LINEに関して1,045件、Slackに関して79件、Discordに関して200件と、計1,324件のツイートが残った。各サービスのツイート内訳を Table 1 に示す。また、抽出したツイートの例を Table 2 に示す。

以降、提案手法におけるフィルタリングおよびベクトル化に関して、機械学習ベースの代替手法との比較を行い、提案手法の有用性を評価する。

### 3.1 フィルタリングの評価

簡便な文字列ベースでのフィルタリングに基づく提案手法に対し、機械学習に基づく置き換えを比較手法とすることで、提案手法の有用性を検証する。まず、提案手法によるフィルタリング前の1,565件のツイートに対し、「検索クエリに用いたサービス名」と「実際にツ

<sup>b)</sup> <https://dash.plotly.com/>

<sup>c)</sup> <https://line.me/ja/>

<sup>d)</sup> <https://slack.com/intl/ja-jp/>

<sup>e)</sup> <https://discord.com/>

<sup>a)</sup> <https://developer.twitter.com/en/docs/twitter-api>

ートの意見が対象としているサービス名」が合致していれば正例、それ以外で負例となるようラベル付けを行った。これを性能評価のためのテストデータとする。次に、比較手法のための学習データとして、2020年1月1日から2020年12月31日の間に投稿されたツイートのうち、「LINE」「機能欲しい」が含まれるツイート1,108件、「Slack」「機能欲しい」が含まれるツイート96件、「Discord」「機能欲しい」が含まれるツイート93件を収集した。各ツイートが真にLINE, Slack, またはDiscordに関するツイートであるかどうかを手動で判定した。各サービスに関して得られた正例、負例の数をTable 3に示す。これを学習データとみなし、各サービスのツイートから単語の特徴ベクトルを抽出し、それぞれを特徴として support vector machine (SVM) を学習した。ベクトル化の方法に関しては bag-of-words (BoW) と TF-IDF を比較する。

提案手法および比較手法により、テストデータを2クラスに分類し、F 値、精度、再現率を算出した。比較手法に関しては異なる学習データにおける評価指標の平均を報告する。各手法によるLINE, Slack, Discord ツイートのフィルタリング結果をそれぞれ Table 4, Table 5, Table 6に示す。総じて提案手法はF 値、再現率に関し

Table 1. Number of tweets regarding each service.

Service name	# of tweets
LINE	1,045
Slack	79
Discord	200

Table 2. Examples of tweets used in the experiment.

Tweet text
LINE にもメッセージにいいね機能欲しい
LINE の誤字脱字とかが多いので、送った後も編集できる機能欲しい
Slack でポスト単位でミュートする機能欲しい
slack でスレッドに関わってる人にメンションできる機能欲しい。@thread 的な
Discord にアンケート機能欲しい
Discord、チャットのポップアウト機能欲しい・・・ (切望)

て比較手法よりも良い結果が得られた。本研究の目指すアイデア抽出は、真に対象サービスに関するアイデアであることと、抽出数の双方が重要である。これを鑑みると、ツイートのフィルタリングによって、正しく対象サービスについて言及しているツイートを抽出でき、かつ正しいツイートを取りこぼさなく抽出できることが求められる。そのため、適合率・再現率の個々の指標ではなく、それらの調和平均である F 値が重要であり、その点で提案手法が最も有効であると考えられる。特に提案手法は、特定の文字列が含まれているツイートを除去するという簡便さにおいても機械学習に基づくフィルタリングに比べて優位と考えられる。

Table 3. The number of positive and negative examples of testing tweets.

	Positive	Negative
LINE	806	302
Slack	54	42
Discord	58	35

Table 4. Results of tweet filtering for LINE.

	F-score	Precision	Recall
Proposed method	<b>0.909</b>	<b>0.834</b>	<b>1.000</b>
TF-IDF+SVM	0.763	0.769	0.791
BoW+SVM	0.777	0.806	0.762

Table 5. Results of tweet filtering for Slack.

	F-score	Precision	Recall
Proposed method	<b>0.784</b>	0.645	<b>1.000</b>
TF-IDF+SVM	0.717	0.639	0.875
BoW+SVM	0.757	<b>0.693</b>	0.856

Table 6. Results of tweet filtering for Discord.

	F-score	Precision	Recall
Proposed method	<b>0.779</b>	0.638	<b>1.000</b>
TF-IDF+SVM	0.751	0.645	0.908
BoW+SVM	0.763	<b>0.665</b>	0.900

### 3.2 ベクトル化手法の評価

提案手法が用いる TF-IDF と、近年の代表的な言語モデルである Bidirectional Encoder Representations from Transformers (BERT)<sup>5)</sup> による文の埋め込みを比較することで、特徴量の有効性を評価する。本研究はツイートのテキストを対象としているため、言語モデルも同ドメインで学習することが望ましい。そこで、大規模なツイート集合で事前学習済みの hottoSNS-BERT<sup>7)</sup> を用いる。ここでは CLS トークンのベクトルと、単語埋め込みの平均ベクトルという 2 通りの方法を比較する。また、提案手法は指定した品詞の単語に基づき TF-IDF を算出するが、全ての単語を用いた TF-IDF も比較する。あわせて、BoW の性能も報告する。

具体的な評価方法としては、まず提案手法によるフィルタリング後のツイート 1,324 件に対して、サービスごとに手動で似た機能のツイートをグルーピングした。その結果、45 個のツイートクラスターと、どのクラスターにも属さないツイート群、各サービスのアイデアではないツイート群に分けられた。45 個のクラスターの情報に基づいて、全てのツイートのペア 875,826 件に対し、同じクラスターに属するかどうかを表す 2 値の正解データを作成した。次に、各手法で算出した特徴ベクトルに基づき全ツイートペア間の類似度を算出し、類似度が上位 5,000 件となるペアを同一アイデアと判定した場合の正解率、精度、再現率、F 値を算出した。Table 7 に各手法の性能を示す。いずれの指標においても提案手法が最も良い数値を示した。BERT を用いると具体的な機能の名称が抽象化されるため、同一アイデアの発見では効果が得られにくいことが示唆された。それに対して提案手法は、ツイート間における特徴的な単語の重複を効果的に検出したものと考えられる。また、全ての単語を用いた TF-IDF は、提案手法である特定の品詞に基づく TF-IDF よりも性能が低下した。これは助詞が文体の影響を受けるためだと考えられる。加えて、BoW に比べて TF-IDF は特徴的な単語を重み付けするため、類似アイデアの発見に適していることがわかる。以上の結果から、アイデアクラスターの発見には提案手法で用いた TF-IDF ベクトルが最も有効であるといえる。

### 3.3 可視化インタフェースの開発

本研究では、提案手法により抽出したアイデアを可視化するインタフェースを開発した。構築したインタフェースの画面を Fig. 1 に示す。ネットワークによる可

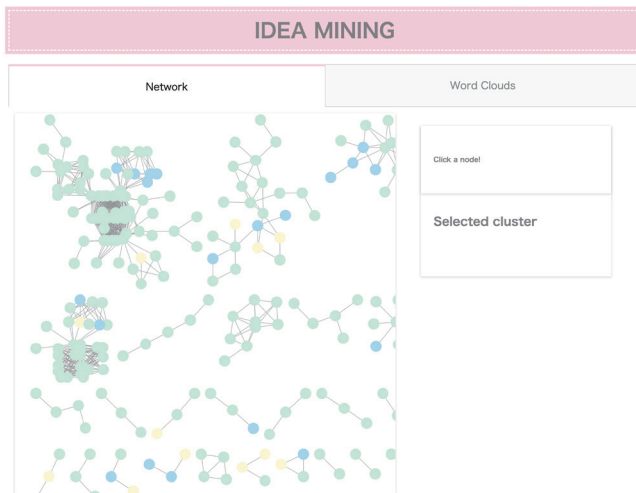
視化とワードクラウドによる可視化の 2 種類を用意し、該当するタブをクリックすることで表示の切り替えを可能とした。ネットワークタブでは、ツイート間のコサイン類似度が閾値 0.6 を超える場合にエッジを構築しており、本実験のデータからは 323 個のノードと、955 個のエッジからなるネットワークが描画された。緑、黄、青で色分けされたノードはそれぞれ LINE, Slack, Discord に関連するツイートに対応する。ノードをクリックした際の画面を Fig. 2 に示す。画面右側には、クリックしたノードのツイート内容と、そのノードが属するクラスターのワードクラウドが表示される。ワードクラウドによって、クラスターを構成するアイデア内容の把握が容易となる。

### 3.4 アイデアの分析

本節では、開発した可視化インタフェースを用いて 3 つのサービスに対するアイデアを分析する。まず、Fig. 1b に示すワードクラウドは、各サービスにおいてどのような機能が多く求められているかの把握につながる。例えば、LINE のワードクラウドでは「いい」「スタンプ」という言葉が大きく表示されており、個々のメッセージに対する『いいね』機能や『スタンプ』機能が多くの LINE ユーザに要望されていることがわかる。次に、Fig. 1a に示すネットワークでは、ワードクラウドで捉えきれないツイートの詳細や、少数意見を確認することができる。クラスターが大きいほど要望の多いアイデアであり、ワードクラウドによる結果と同様、LINE では『いいね』機能に関するアイデアが最も多い結果となった。また、Discord では『メッセージを既読にする』機能のアイデ

Table 7. Results of tweet similarity calculation for clustering ideas.

	Accuracy	Precision	Recall	F-Score
Proposed method	<b>0.952</b>	<b>0.636</b>	<b>0.073</b>	<b>0.131</b>
BERT (CLS token)	0.947	0.202	0.023	0.041
BERT (average)	0.946	0.171	0.019	0.035
TF-IDF (all words)	0.951	0.598	0.068	0.122
BoW	0.950	0.478	0.054	0.098



(a) Home screen.



(b) Word cloud mode.

Fig. 1. Screenshots of the developed interface.

アが見られた。Slackについては、アイデア数は少ないものの、Twitterのボイスチャット機能である『スペース』を要望する声を得られた。最後に、特定のクラスタを詳細に分析する。Fig. 2で示した画面は、「LINEに予約投稿機能が欲しい...」というツイートのノードをクリックした結果である。クリックしたノードが属するクラスタのワードクラウドには、「予約」と「投稿」という単語が出現し、このクラスタが『投稿の予約』機能に関連したアイデアのクラスタであることがわかる。さらにクラスタ内には3つのノード色が存在することから、この機能はどのチャットサービスでも必要とされていると考えられる。以上により、収集したツイートに対してアイデアを分析し可視化することで、各サービスで求められている機能、すなわちサービス改善のためのアイデアを得られることが確認できた。

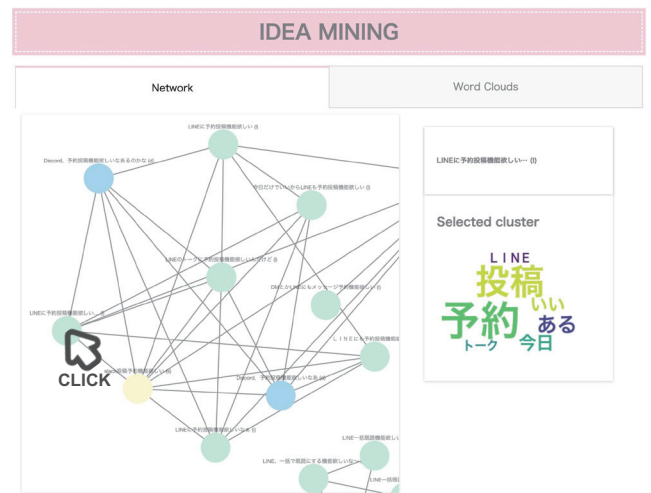


Fig. 2. Screenshot after clicking a specific tweet node.

#### 4 おわりに

本研究では、ツイート分析に基づくサービス改善のためのアイデア抽出と可視化の手法を提案し、結果を分析するためのインターフェースを開発した。実験では、3つのチャットサービスに関して提案手法を適用し、それらに対してユーザがもつアイデアを分析した。ツイートのフィルタリングにおいては、機械学習に基づくフィルタリングよりも提案手法のフィルタリングの方が簡便かつ有効であるという結果が得られた。ベクトル化手法の評価においては、BERTよりもTF-IDFを用いる方が類似アイデアを効果的にクラスタリングできることが示唆された。また、TF-IDFの算出においては、全ての単語から算出するよりも、特定の品詞に該当する単語のみを用いる方が有効であった。さらに、可視化インターフェースを用いることで、各サービスに関する様々なアイデアの直感的理解が可能であることがわかった。

今後はツイート収集方法を改善し、より多くのアイデアを収集することを目指す。提案手法では「機能欲しい」という単語に絞ってツイートを収集していたが、この検索クエリを拡張することと、不要な機能への意見が含まれたツイートも同時収集する手法を検討している。さらに、収集したツイートからアイデアをよりの確に捉えるために、様々な自然言語処理手法を導入する予定である。可視化インターフェースについては、任意のサービスへの意見をリアルタイムに収集・分析するツールの開発を検討する。

## 参考文献

- 1) Twitter, “Twitter announces second quarter 2022 results,” [https://s22.q4cdn.com/826641620/files/doc\\_financials/2022/q2/Final.Q2'22\\_Earnings\\_Release.pdf](https://s22.q4cdn.com/826641620/files/doc_financials/2022/q2/Final.Q2'22_Earnings_Release.pdf), 参照 Oct. 13, 2022.
- 2) B. Lin, N. Cassee, A. Serebrenik, G. Bavota, N. Novielli, and M. Lanza, “Opinion Mining for Software Development: A Systematic Literature Review,” *ACM Trans. Softw. Eng. Methodol.*, **31**[3], 1-41 (2022).
- 3) K. Dave, S. Lawrence, and D. M. Pennock, “Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews,” *Proceedings of the 12th International Conference on World Wide Web, WWW'03*, 519-528 (2003).
- 4) Z. Long, R. Alharthi, and A. E. Saddik, “NeedFull - a Tweet Analysis Platform to Study Human Needs during the COVID-19 Pandemic in New York State,” *IEEE Access*, **8**, 136046-136055 (2020).
- 5) J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, **1**, 4171-4186 (2019).
- 6) Megagon Labs, “Ginza - Japanese NLP Library,” <https://megagonlabs.github.io/ginza/>, 参照 Oct. 13, 2022.
- 7) T. Sakaki, M. Sakae, and N. Gunji, “BERT Pre-trained model Trained on Large-scale Japanese Social Media Corpus,” <https://github.com/hottolink/hottoSNS-bert>, 参照 Oct. 14, 2022.