

テキストマイニング技術の解釈主義的研究への 応用の検討 —Twitter における言説分析を事例に—

小林 洵也^{1,a)}

概要：発表では、Twitter におけるヴィーガン実践者によって発信される投稿を中心としたヴィーガン言説を題材に、テキストマイニング技術を活用してどのように解釈主義的研究を行うことができるのかについて検討する。発表者が現在専攻する分野は文化人類学であり、学部論文では Twitter から特定多数のアカウントによる近年の全投稿と、投稿へのリプライや引用リツイートを自動で収集する手法として、Twitter API と JavaScript を使ったスクレイピングを行った。しかし、その際は形態素解析などの高度な技術は使わずに、最終的には収集したテキストを目で読みナラティブデータとして扱うに留まった。今回はそれを拡張し、より高度なテキストマイニング手法を扱った質的分析と研究対象の量的な抽出法が、エスノグラフィーのような人類学における主要な研究方法論にどのように応用可能であるかについて検討考察する。

キーワード：テキストマイニング，インターネット，Twitter，エスノグラフィー，人類学

An Interpretivist Research Using Text Mining Techniques: A Case Study of Twitter Discourse Analysis

JUNYA KOBAYASHI^{1,a)}

Abstract: In this presentation, I will discuss how text mining techniques can be used to conduct interpretivist research on vegan discourse on Twitter, focusing on posts by vegan practitioners. My current major is cultural anthropology, and in my undergraduate thesis, I used Twitter API and JavaScript for scraping to automatically collect all recent posts from Twitter by a specific number of accounts, as well as replies and quoted retweets to those posts. However, I did not use advanced techniques such as morphological analysis but read the collected text and treated it as narrative data. In this paper, I extend this approach and discuss how qualitative analysis and quantitative extraction of research objects using more advanced text mining techniques can be applied to major research methodologies in anthropology, and ethnography.

Keywords: Text Mining, internet, Twitter, Ethnography, Anthropology

1. はじめに

現代において、インターネットを介したデジタル技術は、日常生活における様々な状況において媒介となって偏在

している。そのため人々の日常で行われている実践を参与観察によって記述し、フィールドにおいて思索する人類学は、デジタル技術をどのような形で研究対象と扱う必要があるのか議論するの必要に迫られている。もはや人類学者でさえ研究協力者とソーシャルメディアなどのインターネットを通じて連絡を取り合い、日常における研究協力者の Twitter でのつぶやきなどを目にする時代において、物理的環境だけに研究の焦点を留める必然性は見つからない。

1 千葉大学大学院人文公共学府
Hitachi Ltd., Chiyoda, Tokyo 101-0062, Japan

a) 21dm1114@student.gs.chiba-u.jp

しかしながらインターネット上の言説空間に広がるビッグデータは手作業では扱えないほど膨大であり、従来の人類学が行ってきた物理的手法だけでは状況によっては手に負えないことも確かである。そのため研究対象となる生活実践者が彼らの生活の一部として営むソーシャルメディアを人類学者が民族誌の一つとして扱うために、情報学におけるテキストマイニング技術がどのように利用可能であるかについて検討する。

2. 事例

本発表で方法的議論を行うために、筆者が過去に行った Twitter における言説分析を題材とする。この研究では Twitter におけるヴィーガン実践者によって発信される投稿を中心としたヴィーガン言説の様態を明らかにするための質的研究であり、一次資料として Tweet データを扱うために収集過程においてテキストマイニング技術を活用した。

この研究においてまず初めに Boellstorff ら[1]のヴァーチャル・エスノグラフィーの具体的手法を参考にした。自身がフィールドに参加しそこで行われる実践を観察する形式を取るために、自分自身も Twitter アカウントを作成し、対象となるヴィーガン実践者と同じように振る舞い、そこで行われている言説の形成を観察した。そして観察過程で #imvegan キャンペーンというトレンドが一時的に広まり、その時に起こったヴィーガンに関する定義論争に着目した。そこで言説の中心となっていた 5 人の Twitter ユーザーを抽出し、#imvegan キャンペーンが始まってから 6 か月間にわたって追跡観察を行った。そうして観察中に閲覧した投稿を一次資料として整理するために、Twitter API と JavaScript を使ったスクレイピングを行った。スクレイピングによって自動収集された情報は Google App Engine で連携された Google Sheets 上に一覧できる形で整理した。集めたデータは投稿の内容と、ユーザーID、Tweet URL、投稿時間、リツイートと「いいね」の数、ハッシュタグとメディアの有無の真偽値であった。どのように解釈主義的研究を行うことができるのかについて検討する。発表者が現在専攻する分野は文化人類学であり、学部論文では Twitter から特定多数のアカウントによる近年の全投稿と、投稿へのリプライや引用リツイートを自動で収集する手法として、Twitter API と JavaScript を使ったスクレイピングを行った。また、これらのデータを構造的に観察するためにも特徴的な投稿や、リプライなどで関連付けられた他の Twitter ユーザーの投稿などを OneNote 上でマッピングしたデータも作成した。最終的に収集したデータは合計 6000 行に及んだ。

以上の事例においては、分析をする以前のデータ収集においてテキストマイニング技術を扱ったが、分析の際には形態素解析などの高度な技術は使わずに、収集したテキストを目で読みナラティブデータとして扱うに留まった。データの収集プロセスの概要は図 1 を参照されたい。

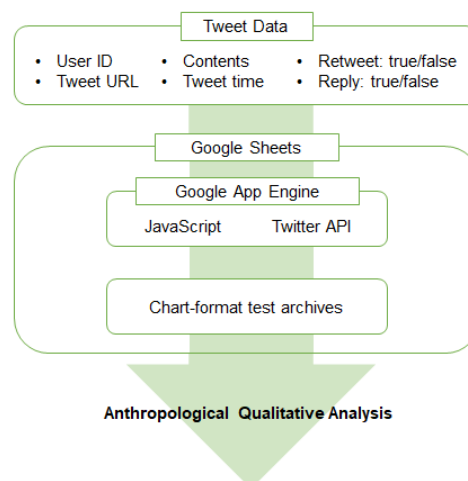


図 1 Tweet データの収集プロセス

Figure 1 Data collection using Twitter API and JavaScript

3. 考察

事例で用いたテキストマイニング手法は、主にデータを整理することを目的としている。後藤[2]は人文情報学の研究傾向として、発見系・解析系・可視化系の 3 つの系列があるとしており、研究過程の内、どこの段階で情報学の技術を扱うかによって大きく分類できるとする。本発表で扱う事例を後藤の提示する分類に当てはめるならば、Twitter から属性ラベルを含む大量のテキスト資料を整理された形で収集する点で、発見系に該当することになる。

また、インターネットを対象にしたエスノグラフィー研究として今回扱った事例を位置づけるならば、木村[3]のサイバーエスノグラフィーの類型が参考となる。木村は、オンラインフィールドワーク、ヴァーチャル・エスノグラフィー、コミュニケーション生態系の 3 つの類型に分け、それぞれ順にインターネット空間を、完結した状態としてのオンライン、オンラインとオフラインのつながり、オンラインを含みこんだ日常、の 3 つのうちどれに焦点を当てるかによって分類できるとした。この場合、今回の事例はオフラインでの実践を考慮に入れていないため、オンラインフィールドワークの類型に属するとみなすことができる。

また発表者はインターネットを対象としたエスノグラフィー研究の場合、関与型と非関与型、関与併用型にも類型化できると考える。近年、エスノグラフィーが人類学のみならず他の研究分野においても方法論として応用が進んでおり、人類学において必要とされている参与観察を伴わずとも、エスノグラフィー研究として扱うようになってきている[4]。今回の事例は研究の主要なトピックとなる事象を選ぶためにまず自身もヴィーガン実践者の Twitter ユーザーとして振る舞い、内側から経験的にトピックとなる事象を抽出

している。一方でデータの抽出・分析においては関与を伴っていないため、今回の事例は関与併用型に分類できると考えられる。

今回の事例を今後より発展的な研究につなげていくためには、テキストマイニング技術をエスノグラフィー研究に方法論的な枠組みとして深く組み込んでいく必要がある。Hsu[5]はエスノグラフィーの特徴として、人間の相互作用や表現、文化的プロセスを解釈し意味を導き出すことを挙げる一方で、コンピュータは情報を量的に扱い再構成することができる点においてスケラビリティという特徴を持つという。そのため計算機技術は、研究者が扱うデータの絞り込みや拡大、および縮尺の変化によって、対象に対するサンプリング方法を再考する機会をもたらすと論じる。そのため解釈的アプローチとしてのエスノグラフィーに新たな解釈の余地、多様な視点を提供するために、エスノグラフィーの分析の部分つまり解析系に該当するようなテキストマイニング技術の組み込みが必要である。

4. おわりに

本発表では、テキストマイニング技術のエスノグラフィーを主とした解釈的研究への応用について事例を通して検討した。現代においてインターネットが日常にさらに広がりを見せる中で、従来の人類学の研究はそれに対応して変容する必要に迫られている。計算機技術の従来への質的研究への組み込みに関して、今後は研究手法の議論のみならず、より抽象的な研究方法論上の議論まで発展させたい。

謝辞 原稿、最終発表と、終始ご指導を戴いた千葉大学人文社会科学系教育研究機構助教小風尚樹先生に感謝する。発表の題材となった学部の卒業論文の時から現在に至るまで、ご指導を戴いている千葉大学大学院国際学術研究院准教授 Ioannis Gaitanidis 先生に感謝する。

参考文献

- [1] Boellstorff, T., Nardi, B., Pearce, C. and Taylor, T. L.: Ethnography and virtual worlds: A handbook of method. Princeton University Press (2012).
- [2] 国立歴史民俗博物館 (編): 歴史研究と〈総合資料学〉, 後藤真: 日本における人文情報学の全体像と総合資料学, pp.1-16, 吉川弘文館 (2018).
- [3] 木村忠正: ハイブリッド・エスノグラフィー — NC 研究の質的方法と実践, pp.121-124, 新曜社 (2018).
- [4] Ingold, T.: That's enough about ethnography!, Journal of Ethnographic Theory, Vol.4, No.1, pp.383-395 (2014).
- [5] Hsu, F. W.: Digital Ethnography Toward Augmented Empiricism: A New Methodological Framework, Journal of Digital Humanities, Vol.3, No.1, pp. 43-64 (2014).