

大学情報系センターへの問い合わせメールの機械的な分類と人手による分類の比較

長島 和平¹⁾, 青山 茂義¹⁾, 根本 貴弘¹⁾

1) 東京農工大学 総合情報メディアセンター

nagashima@go.tuat.ac.jp

Comparison of Mechanical and Manual Classification of Inquiries to University ICT Center

Kazuhei Nagashima¹⁾, Shigeyoshi Aoyama¹⁾, Takahiro Nemoto¹⁾

1) Information Media Center, Tokyo University of Agriculture and Technology.

概要

大学の情報系センターには、提供しているシステム全般に関する多くの問い合わせが寄せられる。メールサービス、学内ネットワーク、ソフトウェア、大学アカウントなど幅広いサービスをサポートするため、窓口で新任の職員や経験の浅い職員が適切な対応をとることが難しい。センター職員による問い合わせ対応を支援するために、適切な対応を提案するシステムが必要であると考えた。そこで、これまでの問い合わせを分類することで、新しい問い合わせが寄せられた際に類似した以前の対応を提案する仕組みの構築を目指す。本稿では、これまでにセンターに寄せられた問い合わせメールについて、機械的にクラスタリングを行った結果と、人手による分類結果の比較を行う。その結果から、機械的な分類手法の問い合わせ窓口への適用可能性を検討する。

1 はじめに

大学の情報系センターでは、研究・教育活動を支援するために多くのサービスを提供している。提供するサービスは多岐にわたり、アカウント、メール、クラウドサービス、学内ネットワークなどが代表的である。また、BYOD(Bring Your Own Device)の導入により、同一の環境下で演習が行えるような仮想デスクトップや無線LANの整備が必要なほか、昨今の感染症対策の影響でリモート講義のためのシステムも導入されている。

情報系センターが提供するサービス範囲は広いものの、予算の制限も大きく、問い合わせに対応する人員は必ずしも潤沢であるとは限らない。本学の場合では窓口で一次対応を行う職員は2~3名である。職員が問い合わせに広く対応するためには、知識と経験が必要になる。専門的な知識を有する職員でも、機関特有の仕組みなどについて理解するための時間が必要である。こうした経験は、ある程度の期間勤務を続けていくうちに身につくものである。

一方で現場では、任期付きの非常勤職員や派遣職員など、短期間の勤務形態での契約も増加している。それに伴い長期間にわたって勤務する職員が減少し、経

験や知識の蓄積の面で問題がある。

一般的に、システムの更新や導入に伴って新たに出てくるものを除けば、ユーザからの問い合わせは過去に同様の事例があることが多い。このとき、過去の対応事例にならって同様の対応を行うことで問い合わせの解決につなげることができる。ただし、経験の浅い職員にとっては問い合わせ内容から過去の事例を探ることが難しいケースもある。センターの業務を円滑に回すためには、経験の浅い職員によるユーザ対応を支援する必要がある。

2 問い合わせの分類における先行研究

先行研究として、山本らによる大学の情報系センターにおけるメールの分類が行われている[1]。山本らの研究では、問い合わせメールの本文をテキストデータに変換し、MeCab^{*1}による形態素解析、Word2Vec[2]による単語のベクトル化とSCDV(Sparse Composite Document Vector)[3]を用いた文書ベクトル化、Kmeans++法[4]を用いたクラスタリングという手法がとられた。クラスタリングの結果、各クラスターでおおむね傾向を抽出することに成功しており、経験の

^{*1} <https://taku910.github.io/mecab/>

ある職員によって得られる結果を再現できたことが報告されており、メールの分類にこの手法を適用することは有用であると考えられる。

一方、クラスタによっては傾向の異なった文書が混在しているケースも確認されており、この問題はクラスタ数を単純に増やすだけでは解決できなかったことが報告されている。こうした課題について、データの前処理の見直しや、本文のみを用いて分類した点について件名を考慮した分析手法の拡張などが今後の課題として挙げられている。

その他に言語処理の観点からも、ある業界で用いられる単語を登録した辞書を作成する研究 [5] や、特に日本語でおこる表記ゆれに対する対策となる研究 [6] もある。情報系センターに寄せられる問い合わせに含まれる固有の用語などについては、事前に辞書登録を行うといった工夫が必要になると考えられる。

3 本研究の目的

本研究では、これまでに寄せられた問い合わせを内容ごとに分類し、新たに問い合わせがあったときに類似した過去の問い合わせとその対応を提示する仕組みの構築を目指す。本稿では、この仕組みの構築に向けて、これまでに寄せられた問い合わせを機械的にクラスタリングしたものと、人手による分類の結果を比較する。

4 問い合わせデータの分類

本稿では、機械的な分類と人手による分類の結果を比較する。機械的な分類手法と結果については、2023年3月に情報処理学会インターネットと運用技術研究会 (IOT 研) において発表した際のデータを利用するため、詳細はそちらの原稿を参照されたい [7]。本章では、対象となるデータと人手による分類手法について述べる。

4.1 分類対象のデータ

本稿における分類対象は、東京農工大学総合情報メディアセンター (以下、「メディアセンター」とする) に寄せられた問い合わせメールのデータを利用する。メディアセンターでは電話、対面、メールでの問い合わせを受け付けているが、分類のためにテキストデータが必要になるため、メールのデータを対象とする。

センター職員によるユーザ対応の支援として、一次対応をスムーズに行うことを目標とする。メール問い合わせにおける一次対応は、最初に送られてきたメールに対する回答にあたる。メール問い合わせにはセン

ター職員側から必要な情報を聞きだすなど数回のやり取りがあることが多いが、一次対応に焦点をあてるため最初に送られたメールのみを見て分類する。

今回対象となるメールは、2022年4月1日から10月31日までにメディアセンターに寄せられた問い合わせで、全部で665件となっている。

4.2 機械的な分類

機械的にデータの分類をするためには、メールのテキストデータをそのまま用いずに前処理を行う必要がある。分かち書きをするうえで、動詞の活用形によるゆれを統一するためにすべて現在形でそろえた。日本語で多く使われるが問い合わせには直接関係のない単語、例えば、問い合わせに必ず含まれてくる「よろしく」などの単語をストップワードに登録している。

それに加え、メディアセンターで扱うサービスに関連する用語を辞書として登録した。例えば本学の場合、提供しているメールサービスが一種類ではなく、全員に発行される Gmail と Microsoft365 のアドレス (それぞれ通称 go メールと me メール)、教職員が役職などで利用するために追加で申請できるアドレス (m2 メール)、m2メールの前身にあたるアドレス (cc メール) がある。「メール」という大枠の分類ではなく、「go メール」や「m2メール」など、提供している各サービスを辞書に登録しておくことで、細かな分類の可能性も考慮した。

このほかに、同一の内容に対する問い合わせであるが表記にゆれがあるものもある。例えば無線 LAN、Wi-Fi、tuatnet*²など、いずれも本学で提供している無線 LAN を指しているが、人によって異なる書き方をしていることがある。これらを同一のものとし、辞書を作成することで、表記ゆれによる違いを吸収した。

以上のような前処理を踏まえ、山本らの手法 [1] にならないクラスタリングした結果の散布図が図 1 である。今回のクラスタ数は目視での確認の結果精度が良いとみられる 7 とした。特にクラスタ 4 の分布範囲が狭く、よく特徴づけられているとみられる。また、クラスタ 3 など広範囲にわたって分布しているものは、幅広い問い合わせが分類されているものと推測される [7]。

4.3 人手による分類の手法

機械的な分類とは異なり、人手による分類では特段の前処理は行わない。分類には最初に届いた問い合わ

*² 本学の無線 LAN の ESSID である

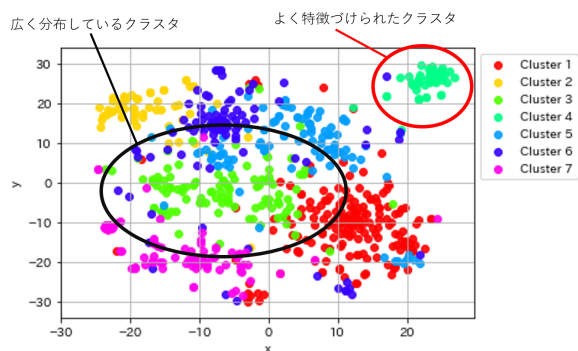


図1 機械による分類結果の散布図

セメールのみを用い、その後のやり取りは含めない。そのため、情報が不足しているような適切な対応が不明な問い合わせがあったとしても、その内容から問題を推測して分類する。

機械による分類ではクラスタを7つに分け、それぞれの特徴量から各クラスタの内容を推測したが、人手による分類では、よくある質問をあらかじめいくつかカテゴライズして、その中に分類する方法を取る。センターに寄せられる質問の内容を鑑みて、人手による分類は以下の10の分類とした。

- 教育研究システム利用
- 電子メール
- ログイン・認証方法
- 有線ネットワーク
- 無線ネットワーク
- VPN
- クラウドサービス (Google, Microsoft 365)
- オンライン授業・会議 (Zoom など)
- アカウント初期設定
- その他

5 分類結果の比較

機械的に分類を行った問い合わせが、人手による分類とどの程度マッチしていたか比較を行う。今回は、665件のメール全ての比較は行わず、機械的な分類でよく特徴づけられていたとみられるクラスタと、散布図で広範囲に広がっていたクラスタに割り当てられていたメールを人手で分類する。

5.1 特徴づけられていたとみられるクラスタ

散布図上で範囲が狭くなっていたクラスタ4の特徴量を図2に示す。特徴量はTF-IDF値を使ってこのクラスタに分類された問い合わせに多く出現する単語

から算出しており、このクラスタでは「容量」「ドライブ」「制限」などの単語が特徴として見られた。Google Workspace 向けのドライブ利用容量が、一つの組織で100TBまでという制限がかかったことに伴う問い合わせが含まれていると予想される。

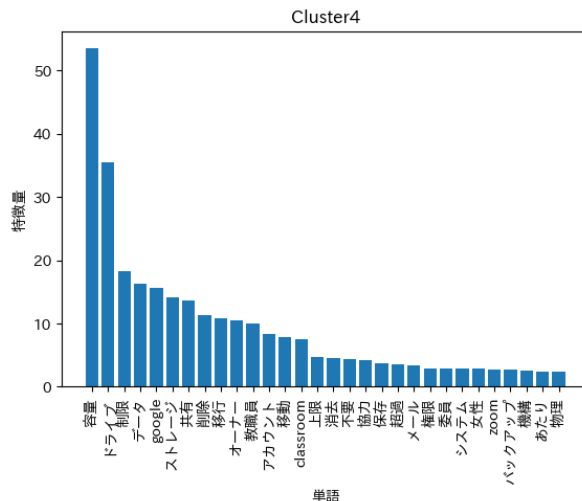


図2 クラスタ4の特徴量

クラスタ4に割り振られていた問い合わせメールに対して、実際に人手による分類を行った。クラスタ4の問い合わせは全部で43件あり、4.3節で示した10カテゴリーのいずれかに分類をさせた。その結果、42件がGoogleやMicrosoft365などを含むクラウドサービスに、1件がその他に分類できた。特徴量からの予想通り、Google Workspaceの容量制限に関する問い合わせが多数を占めており、それ以外にもClassroomに関する問い合わせなど、クラウドサービスに関連するものであった。1件のみ、Webホスティングサービスで提供する研究室Webページの管理者移行に関する問い合わせがこのクラスタに含まれており、「移行」「管理」などの単語から紛れ込んだことが推測される。

5.2 広範囲に広がっていたクラスタ

散布図上では広範囲にわたっていたクラスタ3の特徴量を図3に示す。このクラスタでは、「申請」「接続」「フレンドリー」「システム」などの単語が特徴として見られる。「フレンドリー」は、各ユーザが1つだけ自由に名前を付けられるフレンドリーメールアドレスから抽出された単語とみられる。フレンドリーメールアドレスはじめとした、各種申請についての問い合わせが、このクラスタに属していると推測する。

クラスタ3に割り振られていた問い合わせは全部で107件があった。人手による分類を行ったところ、システム利用が1件、電子メールが18件、ログイン・認

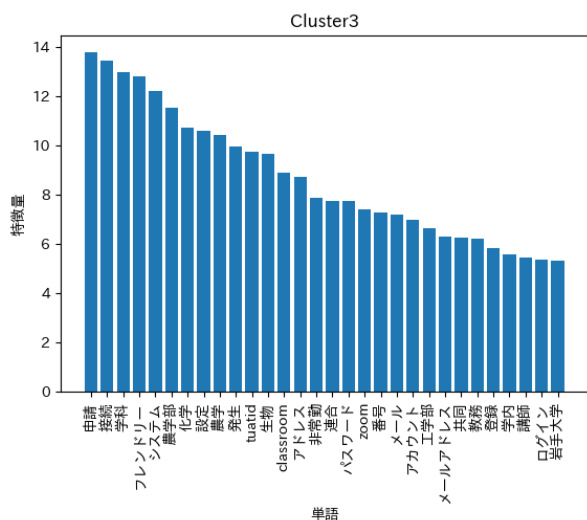


図3 クラスタ3の特徴量

証方法が10件、有線ネットワークが9件、無線ネットワークが13件、VPNが1件、クラウドサービスが13件、オンライン授業・会議ツールが9件、アカウント初期設定が20件、その他が14件となり、カテゴリ上では多くの問い合わせが入り混じった形になった。

ただし、まったく関係のないものが一緒になっているとも言え切れず、例えばZoomに関する問い合わせについては、Zoomへのログインができないことに対してのものが多く、ログインに分類することも可能な内容であった。電子メール、有線ネットワークについては、利用時に申請が必要な点が共通しており、「申請」という観点からは関係の近いものである。

5.3 考察

よく特徴づけられていると考えられたクラスタを人手で分類したところ、予想通りおおむね同じ内容に対する問い合わせが集まっている様子がみられた。集計期間中にGoogle Workspaceの容量制限がかかったことから、同様の問い合わせが多くなっていたことも要因として考えられるが、このクラスタは機械的な分類ができていると考えられる。

一方で、散布図上で広範囲に広がっていたクラスタについては、大枠で類似した問い合わせが含まれている様子も見られたものの、全体としては雑多な種類の問い合わせが集められており、十分な分類がされているとはいえなかった。

仮にGoogle Workspaceのように問い合わせ件数が多いものが分類しやすいのであれば、よくある質問に対しては回答しやすくなる可能性がある。今後ほかのクラスタについても人手による調査を進め、機械的な

分類の可能性を探っていく。

また、今回の人手による分類では提供するサービス内容を元に10のカテゴリを設定したものの、オンライン授業ツールに入れるべきかログインに入れるべきかなど、複数カテゴリにまたがるような問い合わせもあった。例えばZoomのログイン方法に関する問い合わせは、他のシステムと異なるZoom特有の対応が必要になるため、今回の人手による分類ではオンライン授業ツールへと分類したものの、問い合わせ内容としては「ログインできない」ということになる。人手による分類手法についても、過不足なくカテゴリ化できるように検討が必要である。

6 まとめ

大学の情報系センターに届いた問い合わせメールについて、機械的な分類の結果と人手による分類の結果を比較し、機械的な分類の有用性を調査した。その結果、散布図上で範囲が狭いクラスタに分類された問い合わせは、人手による確認でも同様の内容が分類されていることが確認できた。散布図上で範囲が広がっているクラスタでは、様々な問い合わせが含まれていることが確認でき、効果的な分類がなされていなかった。これらの結果から、全体としては窓口対応への応用のレベルには達していないものの、よく特徴づけられたクラスタの活用は現時点でも可能性があると考えられる。

今後は、よく特徴づけられたクラスタの窓口対応への応用の可能性と、全体的な精度の向上を目指す。また、今回は仮で設定した人手による分類のカテゴリについても、サービスベース、対応内容ベースなど多角的に検討していく。

参考文献

- [1] 山本 一幸, 大瀧 保広, 佐藤 伸也, 畠田 敏行, 野口 宏, 羽瀧 裕真, 外岡 秀行, 問合せデータの分散表現を用いた分類, 学術情報処理研究, 24, 1, 68-77, 2020.
- [2] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, Jeff Dean, Distributed Representations of Words and Phrases and their Compositionality, Advances in Neural Information Processing Systems, 26, 2013.
- [3] Dheeraj Mekala, Vivek Gupta, Bhargavi Paranjape, Harish Karnick, SCDV : Sparse Composite Document Vectors using soft clustering over

- distributional representations, Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 659-669, 2016.
- [4] David Arthur, Sergei Vassilvitskii, K-Means++: The Advantages of Careful Seeding, Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 1027-1035, 2007.
- [5] 相良 かおる, 小野 正子, 実践医療用語辞書 ComeJisyoSjis-1 の作成, 言語処理学会第 25 回年次大会 発表論文集, 2019.
- [6] 岡部 浩司, 河原 大輔, 黒橋 禎夫, 代表表記による自然言語リソースの整備, 言語処理学会第 13 回年次大会, 2007.
- [7] 長島 和平, 青山 茂義, 根本 貴弘, 三島 和宏, 大学情報系センターへの問い合わせメールの分類手法の検討, 研究報告インターネットと運用技術 (IOT), 2023-IOT-60, 10, 1-6, 2023.