

モバイルアプリ上での機械学習による情報整理と検索性向上の試み Improve searchability through machine learning on mobile applications

嶺田 あんず¹⁾ 田中 康一郎¹⁾
Anzu Mineda Koichiro Tanaka

1 はじめに

スマートフォンアプリが日常生活に深く浸透する中で、膨大な情報の中から必要な情報を効率よく取得できるような設計が求められている。とりわけ、情報の整理や提示の仕方がユーザビリティに直結するようなアプリでは、機械学習を活用してテキストを自動的に分類・整理する技術が、有効な解決方法として注目を集めている。本研究では、九州産業大学公式アプリ「KSU」[1]の開発と運用を継続して行っており、このアプリを活用した多くの研究も行われてきた[2]。本研究では、図1、図2に示すこのアプリのお知らせ機能に機械学習による自動分類機能を導入し、お知らせ情報の整理と検索性向上を目指したシステムを設計・実装することを目的とする。自動分類機能を構築した結果、導入後には一定の分類精度の向上が確認された[3]。ただし、学習データが150件を超えた段階で精度(Accuracy)の伸びが鈍化する傾向も見られた。

本研究では、この精度向上の鈍化要因の一つとして考えられる「カテゴリ間のデータ数の偏り」に着目し、データ数を均等に調整した上で再評価を行った。また、アプリの運用フェーズを見据え、分類精度の向上を目的としてユーザから提供された誤分類情報をもとに、モデルを継続的に学習・更新できる仕組みを設計・実装した。これにより、分類性能の向上だけでなく、アプリの柔軟性や将来的な拡張性を高めることを目指す。

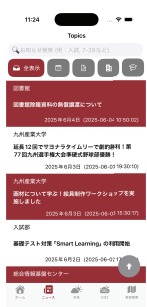


図1 全表示



図2 入試



図3 誤分類報告

2 開発環境

本研究では、iOS版「KSU」アプリにおける通知分類機能の開発に、Appleの統合開発環境Xcode(16.4)およびSwift(6.1.2)を使用した。機械学習モデルの作成にはCreate MLを使用し、データ前処理からモデル学習・評価・エクスポートまでのプロセスを自動化スクリプトを開発した。学習済みモデルはCore ML形式(.mlmodel)として出力され、サーバ上に配置している。アプリはこのモデルにアクセスして分類処理を行うため、端末のリソースに依存せず柔軟な運用が可能である。

1) 九州産業大学理工学部情報科学科. Department of Information Science, Kyushu Sangyo University.

サーバサイドの実装には、PHP(8.1.32)とMariaDB(10.5.27)を用い、誤分類報告を受け取ってJSON形式で保存する仕組みを構築した。また、macOS環境の開発用サーバでは、収集データをもとに機械学習モデルの再学習を行う自動処理を実装しており、モデル更新を円滑に行えるよう設計している。

3 アプリの仕様

「KSU」アプリは、九州産業大学の学生や教職員を主な対象とするアプリである。アプリの機能の一つである「お知らせ」機能では、大学の公式サイト11ヶ所からお知らせ情報をJSON形式で取得し、タイトル一覧として表示する仕組みを備えている。

各お知らせをタップすると、該当するWebページがブラウザで開き、詳細を確認できる。従来の一覧表示に加え、お知らせの内容に応じて「行事」「報告」「就活」「入試」のカテゴリに自動分類される機能が導入された。

4 システムの仕様

本システムは、「アプリ」「中継サーバ」「再学習サーバ」の3層構造で構成される。アプリでは、ユーザが通知内容に対する誤分類を報告できるUI(図3)を備えており、誤分類の報告は中継サーバへ送信される。中継サーバでは、報告内容を蓄積・集計し、定期的にJSON形式のデータセットとして整理される。

一定数以上の報告が蓄積された場合や、定期的なスケジュールに基づき、再学習サーバがこのデータセットを取り込み、Create MLを用いた自動再学習処理を実行する。再学習後のモデルはCore ML形式で再出力され、サーバ上に再配置される。

5 実装

本研究では、学習データ数の増加に伴う分類精度向上の鈍化要因として、カテゴリ間のデータ数の偏りが影響している可能性に着目した。これを検証するため、「行事」「報告」「就活」「入試」の4カテゴリについて、それぞれ50件ずつ均等に構成したデータセット(計200件)を作成し、その条件下で分類モデルの再学習と評価を実施した(表1)。元データにはカテゴリごとに件数の偏りがあるため、均等条件で使用可能な最大件数は200件であった。

表1 データの内訳

	データ数	行事	報告	就活	入試
実データ	200個	76個	98個	10個	18個
均等調整	200個	50個	50個	50個	50個

さらに、分類モデルの継続的な精度改善を目的として、実際の利用者から得られる誤分類データを活用する再学習システムを実装した。図4にシステム全体の構成を示す。アプリでは、各通知に対して表示されたカテゴリが適切かをユーザが確認でき、誤分類だと感じた場合には正しいカテゴリを選択して報告するUI(図3)が提

供される。ユーザが送信した誤分類報告は、JSON形式で中継サーバに蓄積される。

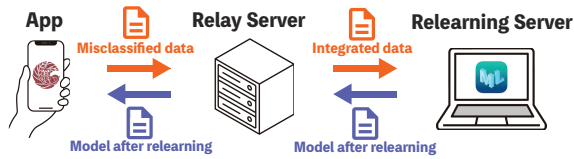


図4 システム構成

中継サーバはPHPで構築され、報告データの受信と保存、そして一定時間ごとの統合処理を担う。誤分類データは一時領域に保持され、1時間ごとに主データセットへ統合される仕組みとなっている。統合処理では、学習済みデータとの重複チェックやバックアップ作成が実施され、処理済みデータは削除されることで、二重学習を防止している。

再学習サーバはmacOS環境上に構築され、Swiftスクリプトを用いて再学習処理を自動化している。このスクリプトはAppleが提供するCreate MLのCLIを用いており、1時間ごとに中継サーバから統合済みデータセットを取得し、分類モデルを再構築する。再学習後のモデルはCore ML形式(.mlmodel)でエクスポートされ、中継サーバに再配置される。アプリは常に最新のモデルで推論するため、ユーザに明示的なアップデート操作を求めることなく分類モデルの更新が可能となっている。

このような構成により、ユーザフィードバックに基づく動的なモデル改善が実現され、精度向上に加えてアプリケーションの柔軟性と将来的な拡張性が確保されている。また、すべての処理がローカルまたはサーバ間で完結するよう設計されており、プライバシー保護やセキュリティの観点からも安全性が担保されている。

6 結果と考察

まず、カテゴリごとのデータ件数を均等にした200件のデータセットを用いた再学習モデルの評価結果では、正解率が偏りのある元データでの約74%から64%へと低下した(図5)。この結果は、もともと学習データ数が多かった「行事」や「報告」のカテゴリで識別性能が低下し、全体精度に影響を与えたことを示している。一方、データ数が少なかった「就活」や「入試」カテゴリでは、学習機会が増えたことで識別性能が改善された可能性も考えられる。

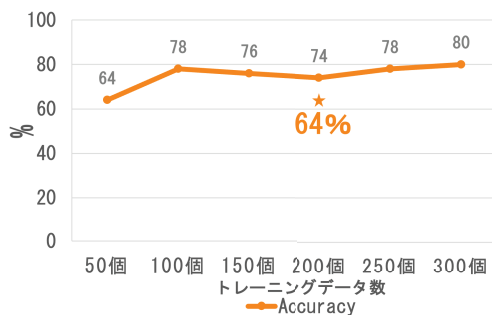


図5 正解率の比較

これより学習データの分布が実際のデータの出現傾向と大きく異なる場合、モデルの性能や汎化能力に影響を

与えることが示唆された。つまり理想的なデータの均等化よりも、実環境に即したデータ構成の方が現実的かつ実用的な分類モデルの構築に適していることが分かる。

次に、誤分類報告を活用した再学習システムの導入は、モデルの継続的な改善において有効な手段であることが確認できた。特に、運用中の実データを即座に学習データに反映できる体制は、モデルの精度維持や適応性の向上に寄与する。また、ユーザ参加型のフィードバック収集により、開発者が想定していなかった誤分類パターンの検出や、学習対象の多様化が期待される。

今後の課題としては、ユーザから収集される誤分類データの質と量の確保、再学習スケジュールの最適化、誤報を含む報告のフィルタリングなどが挙げられる。誤分類報告の信頼性を高めるためには、全ユーザに開放せず、教職員や一定のレビュー権限を持つユーザなど、信頼性の高い限られたユーザにのみ報告機能を提供する運用も有効である。これにより、悪意ある報告やノイズの多いデータが再学習に混入するリスクを低減できる。

また、カテゴリ間の不均衡を解消しつつ、実データの分布も考慮するハイブリッドな学習戦略の導入が求められる。さらに、ユーザからのフィードバックがどの程度分類性能の改善に寄与しているかを定量的に評価し、フィードバックループの効果を可視化することで、より効果的な運用体制の構築を目指す必要がある。

7 まとめ

本研究では九州産業大学公式アプリ「KSU」における通知情報の利便性向上を目的とし、機械学習を用いたお知らせのカテゴリ自動分類機能を実装した。Create MLによって構築したText ClassifierモデルをiOSアプリに組み込み、通知を「行事」「報告」「就活」「入試」の4カテゴリに分類する仕組みを構築した。アプリ上では、分類結果をもとにカテゴリ別に通知が表示され、ユーザは目的の情報に迅速にアクセスできる。また分類処理はアプリ内でリアルタイムに行われ、新しいお知らせにも即時対応できる柔軟な仕様を実現した。カテゴリごとのデータ数の偏りが精度に与える影響を評価するため、各カテゴリ50件・計200件の均等データセットで再学習を行ったが、精度は低下し、実際のデータ分布を考慮した学習の重要性が示唆された。加えて運用後のモデル改善を可能にするため、ユーザが分類ミスを報告し、その情報をもとにサーバ側でモデルを再学習できる仕組みを構築した。これにより、アプリリリース後も精度向上を継続的に図れる実運用に適したシステムを実現した。今後は分類カテゴリの拡張や他のモデルの比較、および精度向上に向けたフィードバック処理の自動化を進めることで、より高いユーザビリティの実現を目指す。

参考文献

- [1] 田中康一郎：学生が創る大学公認アプリ「KSU」、大学時報, No. 392, pp. 54 – 57 (2020).
- [2] 嶺田あんず, 安部萌, 田中康一郎：異なるプログラミング言語によるスマートフォンアプリの実行効率比較, 2024年度第32回電子情報通信学会九州支部学生会講演会, p. D-46 (2024).
- [3] 嶺田あんず, 田中康一郎：スマートフォンアプリにおけるテキスト分類機能の設計と実装, 情報処理学会第87回全国大会, pp. 3-167 – 3-168 (2025).