

小口研究室 研究紹介 (2020年度)

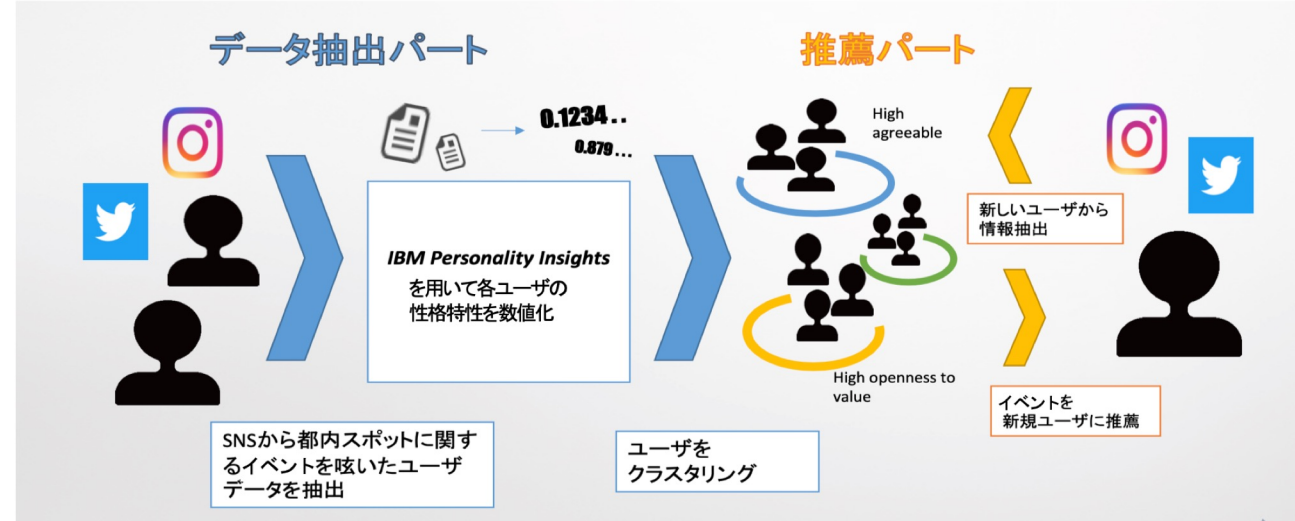
(お茶の水女子大学理学部情報科学科)

ソーシャルメディアにおける性格特性を利用したイベント種別傾向に関する一検討 (研究担当: 伊藤 桃)

研究背景

- ◆ コロナウイルス収束後都内のイベント参加者が急激に増加することが見込まれている
 - ◆ イベント推薦システムはすでに多く存在するが、ユーザの趣味趣向からイベントを推薦するシステムが主流である
 - ◆ より新規性のある推薦システムの需要の高まっている
- ソーシャルメディアから得られるユーザの性格特性を用いたイベント推薦システムの提案

提案システム



- ◆ **データ抽出パート**
 1. 都内イベントに関するツイートを抽出したユーザデータを抽出
 2. 抽出データをPersonality Insightsにて数値に変換
 3. 2.のデータによってユーザをクラスタリング
- ◆ **推薦パート**
 1. 新規ユーザの過去ツイートデータを抽出
 2. 抽出データをPersonality Insightsにて数値に変換
 3. データ抽出パートにて作成したモデルを基にクラスタリンググループのよく行くイベントを推薦

Personality Insights

- ◆ テキスト情報から性格分析をするIBMのサービス
- ◆ 心理学的な性格の大分類とされる、ビッグファイブや人間の欲求を分類した、ニーズ、個人が大切にしている価値観を分類した、価値の3軸52項目をテキストから0-1の連続値に数値化

実験目的

- ◆ 提案システムを実現するため、そもそもイベントごとに参加ユーザ性格値に傾向があるかどうか実験

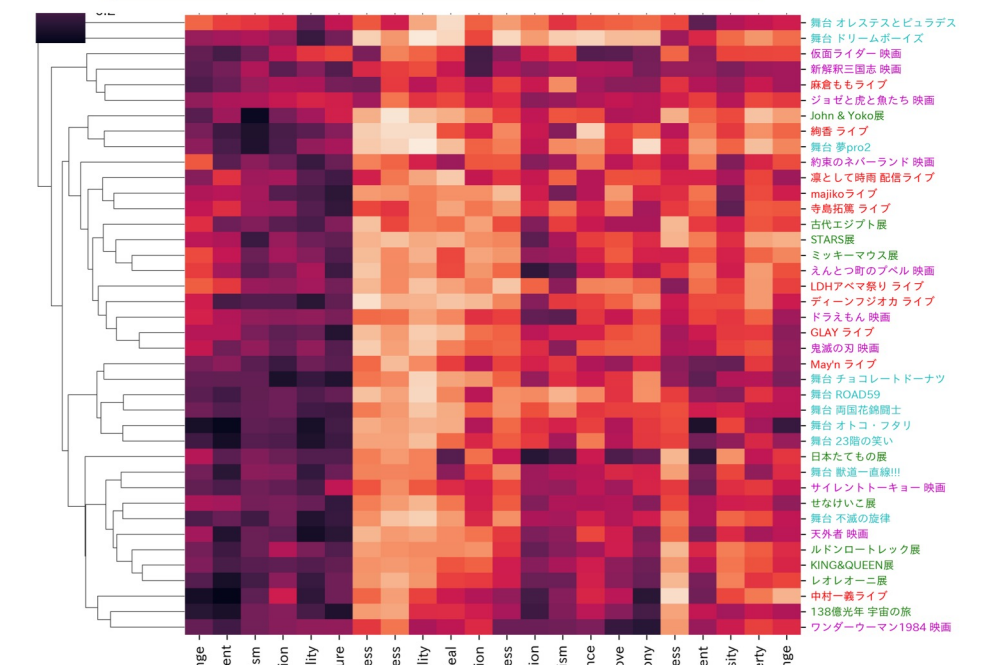
実験概要

- ◆ イベントカテゴリを先行研究から「映画」「舞台」「展示」「ライブ」と定める
- ◆ 各カテゴリごとに10個のイベントを設定
- ◆ 2020年12月30日にTwitterで各イベント名を含むツイートで咬いた日本人ユーザの過去ツイートをテキストデータとしてPIにて性格成分を取得

まとめ、今後の課題

- ◆ イベント種別の性格特性傾向を、感覚的に腑に落ちる結果として確認した
- ◆ 具体的なシステム実現に向けて話を発展させていきたい

実験結果

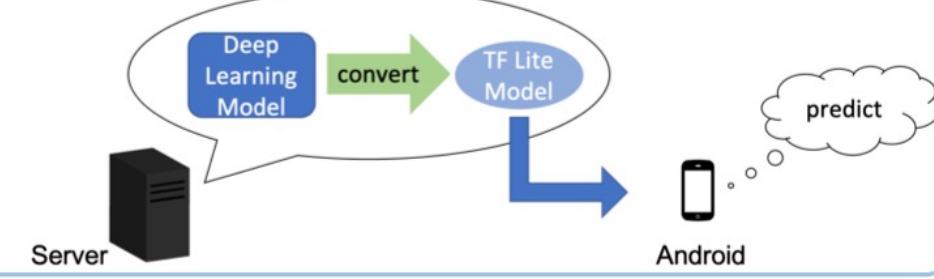


- ◆ 水色: 「舞台」、紫色: 「映画」、緑色: 「美術」、赤色: 「ライブ」
- ◆ 「映画」以外イベントカテゴリごとに固まってクラスタリングされている → 「映画」は種類によって見る層に違いが見られることが考えられる
- ◆ 「美術」カテゴリは全体的にbig5_openness(知的好奇心度)が高い

深層学習を活用したAndroid端末上での時系列データ予測 (研究担当: 佐藤 里香)

研究概要

- ◆ **研究背景**
 - 無線環境下でのトラフィックの輻輳は突発的に生じ、一度起こると制御が難しい上制御を試みるとさらに悪化 → 輻輳を事前予測することが理想
 - 予測にあたりデータを端末外に出す安全上の問題やデータ転送に要する時間等が課題に → 端末内処理が理想
- ◆ **研究方針**
 - サーバで深層学習を用いたトラフィックデータの解析、学習モデルを作成
 - モデルを形式変換し、Androidアプリケーションとして端末に組み込む



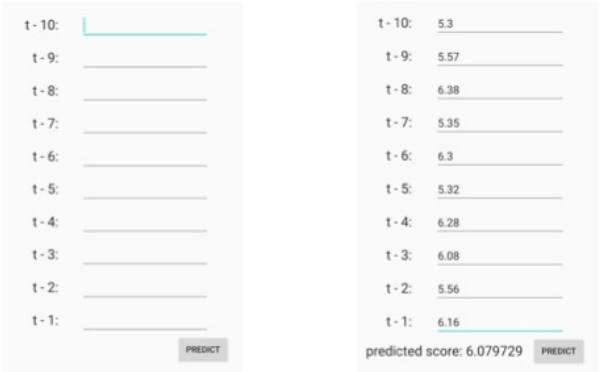
サーバでの学習モデルの作成・形式変換

- ◆ 端末での無線LAN通信のスループットデータを重回帰分析で予測
 - TensorFlowのKerasを使用
 - 端末5台分、181秒間同時通信時のデータを使用
 - 各端末でt-10秒 ~ t-1秒の10秒間のデータからt秒における値を予測
 - データをトレーニングデータ7割、バリデーションデータ3割に分割
 - epoch数300で実験
- ◆ 作成した学習モデルをAndroid端末に導入するため、TensorFlow Lite対応形式に変換
 - 変換前の学習モデルによる予測結果と変換後の学習モデルによる予測結果を比較 → 変換前後で最大でも10⁻⁶程度の誤差に留まっていることが確認できた

変換前のモデル (TF モデル)	変換後のモデル (TF Lite モデル)
0.4860211611	0.4860211611
-0.4430246353	-0.4430246353
5.0874166489	5.0874166489
6.1870837212	6.1870837212
1.2597573996	1.2597573996
-0.2283095270	-0.2283095270
-0.7754652900	-0.7754652900
-0.5668531060	-0.5668531060
-0.5548511147	-0.5548511147
-0.5718665123	-0.5718665123

学習モデルのAndroid端末上への導入・性能評価

- ◆ 形式変換後の学習モデルを組み込んだAndroidアプリケーションを作成
 - t-10秒 ~ t-1秒の10秒間のデータを入力しt秒における値を予測
- ◆ 予測精度の比較
 - サーバ上と端末上の予測結果を比較 → 最大でも10⁻⁶程度の誤差に留まっており、端末上でもサーバとほぼ同等の精度であることが確認できた



サーバ上での予測値	端末上での予測値
0.4860211611	0.4860211611
-0.4430246353	-0.4430246353
5.0874166489	5.0874166489
6.1870837212	6.1870837212
1.2597573996	1.2597573996
-0.2283095270	-0.2283095270
-0.7754652900	-0.7754652900
-0.5668531060	-0.5668531060
-0.5548511147	-0.5548511147
-0.5718665123	-0.5718665123

- ◆ 予測時間の比較
 - サーバ上と端末上の予測時間を比較 → 端末上ではサーバに比べ3倍程度要しているが、想定する輻輳制御周期を考えると十分な速度
- ◆ CPU使用率の比較
 - 端末上での予測処理時のCPU使用率を計測 → TensorFlow Lite公式サイトが提供しているサンプル機械学習アプリに比べ、十分低い値

	サーバ上	端末上
予測時間	1.96ms	6.26ms

	検証用アプリ	サンプルアプリ
CPU使用率	5%	15%

今後の課題

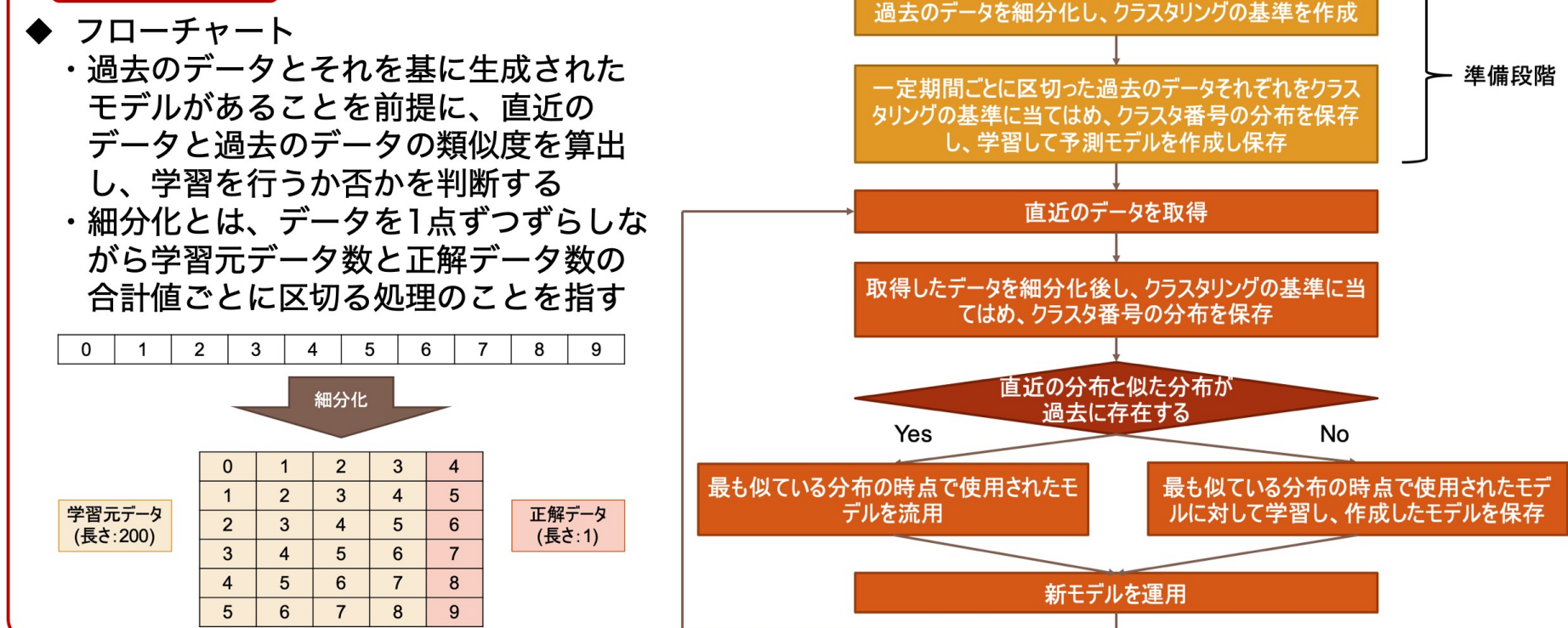
- 機械学習予測の精度向上
- 想定している輻輳制御システムの環境により近い状況のアプリケーション実装とその性能評価

時系列予測モデル学習時間及び計算資源削減手法の検討 (研究担当: 高橋 佑里子)

研究概要

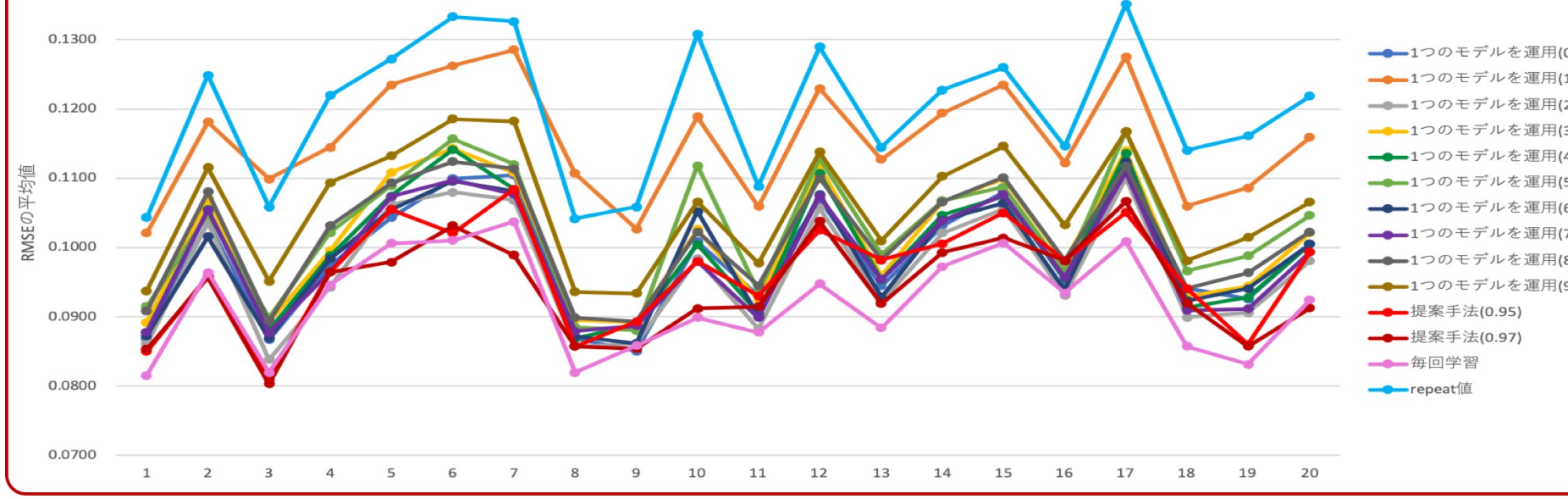
- ◆ **背景**
 - 仮想環境において、計算資源のオーバーコミットに由来するVMの性能低下を防ぐことを目的として、VMのCPU使用率を予測し、その結果に基づいて制御を行う技術が知られている
 - VMとそこで実行されるアプリケーションは時々刻々と変化するため、環境の変化に合わせて予測モデルを継続的に学習し、モデルを更新することで予測精度を担保する
 - しかし、大規模環境におけるモデルの継続的な学習は、多大な学習時間及び計算資源を必要とするため困難
- ◆ **目的**
 - 環境の変化を監視/評価し、精度担保に必要と判断された場合のみ予測モデルの学習/更新を行うことで、学習時間及び計算資源の削減を実現する手法の検討

提案手法



実験

- ◆ **内容**
 - Microsoft社が提供しているAzureのVMデータセットの一部を使用し、100種類の波形を抽出した後、それらを異なる割合で含むデータを30種類作成し、10種類を準備段階、20種類を運用段階として実験
 - 提案手法における「似た分布」の基準は、分布のコサイン類似度が0.95または0.97以上とした
 - 基準を0.95とした場合の学習回数は5回、基準を0.97とした場合の学習時間は13回となった
- ◆ **結果**
 - 提案手法2種類(赤色)の結果は、repeat値や1つのモデルを運用した場合より、おおよその区間で高精度 → 時系列データを細分化したものを同一のクラスタリング基準に当てはめ、得られたクラスタ番号分布の類似度を元に学習を行うか否かを判断することで、学習時間を大幅に削減しながら予測モデルの精度を維持することが可能であることが確認できた



今後の課題

- より良いフローチャートの作成
- ニューラルネットワークのアーキテクチャを最適化する手法の検討