

大規模言語モデルを用いたマイクロブログからの イベント情報抽出に基づく人口変動予測

増田 颯天[†] 豊田 正史^{††} 吉永 直樹^{††}

[†] 東京大学大学院情報理工学系研究科 〒113-8656 東京都文京区本郷 7-3-1

^{††} 東京大学生産技術研究所 〒113-8505 東京都目黒区駒場 4-6-1

E-mail: †{hmasuda,toyoda,ynaga}@tkl.iis.u-tokyo.ac.jp

あらまし イベント会場周辺の人口変動は、その日のイベントの有無や種類によって大きく変化し、周囲に様々な影響を与えるため、その予測は重要な課題である。イベント会場における人口の予測には将来のイベント情報が必要となるが、過去の人口時系列データのみでは精度の高い予測が難しく、それに対してイベント会場と予測する日付に言及したマイクロブログ投稿の集合を入力に加えることで精度を改善する手法が提案されている。しかし、投稿に含まれる情報は多様であり、中には実際のイベントとは無関係なノイズとなるような投稿も含まれる問題や、様々なアーティストが利用するコンサート会場では予測精度が低いといった課題がある。本研究では、内容が多様であるマイクロブログ投稿から適切に特徴量を抽出するとともに結果の解釈性を向上させるために、大規模言語モデルによりマイクロブログ投稿集合からイベント情報を抽出し人口変動の時系列予測モデルに用いる手法を提案する。実験では混雑統計データと Twitter データセットを用いて関東のイベント会場における 24 時間先までの人口予測を行い、その精度を調査した。

キーワード ソーシャルメディア, 地理データ, 行動データ, LLM

1 はじめに

野球の試合やコンサートなどのイベントが発生するとき、発生場所やその周辺に参加者が集まることで、その地域の通常時以上の人口が突発的に密集することになる。このような人口の急増は、イベント参加者だけでなくその周辺地域や沿線沿いの人々にも悪影響を及ぼす可能性がある。たとえば、周辺地域においてゴミ箱以外に大量のごみが捨てられたりトイレの量が不足したりなどの衛生的な影響や、電車やバスの混雑、道路の渋滞など交通の影響などである。

イベントによる人口の変動を予測することは、こうした問題に対処する上で重要である。事前にイベントの発生する場所や時間、人口を予測することで、ゴミの回収の強化や交通整理の人員の投入などを計画することが可能となり、また非イベント参加者には混雑地域を迂回する移動ルートの提案なども可能となる。しかしながら、イベント会場のような不定期的なイベントにより人口が周期的に変動しない場所においては、駅などの周期的に人口が変動する場所と比較して予測誤差が大きいことがわかっている [1]。これは過去の人口の時系列データのみでの入力では、将来に起こる不定期的な大規模イベントの有無や種類の判別が困難であることに起因するものである。この問題に対処するための研究として、人口データとは異なるデータを入力に使用する研究が行われている。具体的には、地図アプリの検索クエリを使用した研究 [2] や、乗り換え案内アプリのログを用いた研究 [3], Twitter (現 X) を代表とするマイクロブログ上の投稿を用いた研究 [4] などが挙げられる。とりわけ、マイ

クロブログの投稿は多様なユーザーが様々なトピックについて日々投稿しているため、その大量の投稿の中に将来のイベントに言及した投稿も含まれており、イベントの発生場所や時間などへの有効な手掛かりとなる。しかし、こうした投稿中にはイベントと無関係な情報などノイズとなるような投稿も含まれていたり、また、マイクロブログ投稿は文の形式が一定でなく曖昧な表現で書かれることも多く、これらの問題が予測精度を下げる要因であり、また投稿群を見た際に予測出力の根拠となったイベント情報がわかりづらくなる原因でもある。

本研究では、ノイズを含むマイクロブログ投稿からイベント情報を抽出する手法として大規模言語モデル (Large Language Model, LLM) を活用する手法を提案する。投稿の集合を LLM に入力し一定の形式に沿って投稿に含まれる未来のイベント情報を要約し、その要約を人口時系列予測への追加入力として扱う。また、LLM を用いてイベント情報をテキストとして抽出することにより結果の解釈性も同時に向上させることができる。

実験では、イベント会場として関東圏の 8 つの会場を対象として選択し、予測タスクにおいて時系列予測の予測誤差の精度とイベントの有無の分類精度を評価した。既存手法との比較のためベースラインとして Tsukada ら [5] の手法を用い、データの実験期間として 2014 年 12 月から 2018 年 11 月までの 4 年間を設定し、マイクロブログのデータとして研究室で収集している Twitter (現 X) のデータ [5] を使用した。結果として、提案手法は 8 会場中 7 会場で従来の手法と同等以上の精度を達成するとともに、イベント情報の要約により結果の解釈性が向上した。

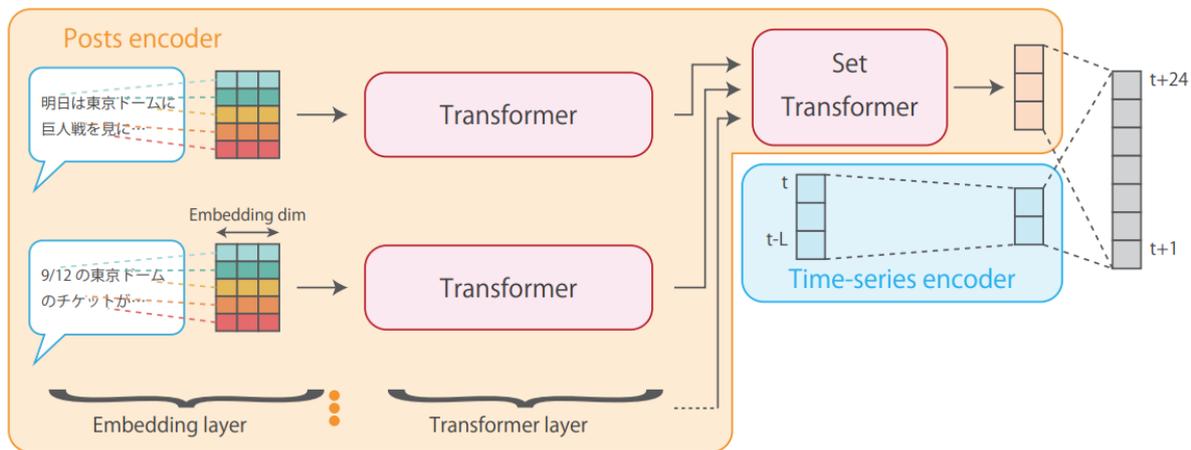


図1 Tsukada らが提案した SetTransformer を用いたモデルの概要 [5].

2 関連研究

イベント会場周辺での大規模イベントによる不規則的な人口や交通の変動を予測するために、様々なデータを活用する研究が行われている。Liao ら [2] は、イベントの発生前にイベント会場に対する地図検索クエリが増加することに着目し地図アプリのオンライン検索クエリを人口予測に用いており、Konishi ら [3] は同様に乗り換え案内アプリの検索クエリを活用した手法を提案している。本節では、マイクロブログを活用した人口変動予測の研究を紹介する。まず、Tsukada ら [4] による、マイクロブログのイベント情報を用いた非周期的な人口変動の予測手法を取り上げる。次に、Liang ら [6] の研究を紹介し、移動需要予測における大規模言語モデル (LLM) の有用性を示す。

2.1 マイクロブログ投稿を用いた人口予測

Tsukada ら [1], [4], [5] は、イベント会場における予測タスクにおいて、入力として人口変動の時系列データとイベントに関連したマイクロブログ上の投稿群のデータ融合を行い、また投稿の数や単語の出現頻度だけでなく投稿の意味的な内容を捉えるために Self Attention ベースのモジュールを組み込んだモデルを提案した。

収集されたマイクロブログ投稿の集合を入力として扱う際の課題は、集合内の投稿の数が予測日ごとに変化することであり、これは set-input 問題として扱われる。set-input 問題では、集合内の要素の順番によらず出力が一定である順序不変性も要求され、これに対応するために Tsukada らは Set Transformer [7] をモジュールとして組み込んだモデルを提案し、その精度を時系列データのみを入力としたモデルとの比較を行った。実験では、入力にマイクロブログ投稿を追加した提案手法が、時系列データのみを入力としたモデルに対して精度の優位性が示されており、実際にマイクロブログ投稿が将来のイベント情報の手掛かりとして有効であることが示されている。

しかし、これらの研究においては残された課題として、予測の根拠となる投稿を提示することなどによる結果の解釈の方法の検討や、イベント日数やマイクロブログ投稿の少ない会場に

おける手法の適用などが挙げられている。この課題に対する方針として、次の LLM を用いた研究を紹介する。

2.2 LLM を用いた移動予測

Liang ら [6] は、LLM を活用して公共イベントにおける人の移動需要を予測する LLM-MPE フレームワークを提案している。この手法では、大規模言語モデルに対してプロンプトとしてイベント情報と過去の実際の移動需要を入力することで、将来の需要を LLM に予測を行わせる。この手法は、イベント会場の公式サイトより収集したイベント情報のテキストを LLM によりイベント特徴量として文章の標準化を行う段階と、過去数日分の標準化したイベント特徴量と実際の移動需要をプロンプトとして LLM に与え次の 1 日の移動需要を予測させる段階からなる。こうしたタスクに LLM を用いる利点は、LLM が広範なドメインのウェブテキストにより訓練されているため様々な知識と言語能力を獲得している点であり、実際にイベント情報の標準化において、バスケットボールチームから音楽歌手までの様々な固有名詞に基づいて欠落したイベント情報を補完し、また不要な詳細情報を要約する能力があることが報告されている。

一方で、この手法は過去数日分の入力から直後の 1 日分の移動需要を予測する手法であり、その 1 日の中で時間帯ごとによいように需要が変化するかという複数ステップでの予測はできない。また、LLM の fine-tuning を行っておらず、プロンプトに過去数日分の情報のみを入力するため、予測時に訓練期間全体の情報をモデルに与えることができない。本研究では、LLM によるイベント情報の要約の手法とマイクロブログを用いた人口の時系列予測の研究を融合し、イベント会場における非定期的な人口変化の長期予測を行う。

3 提案手法

本研究では、マイクロブログ投稿を用いたイベント会場周辺の非定期的な人口変動の時系列予測タスクにおいて、投稿群のノイズ除去と結果の解釈性の向上のために大規模言語モデルによるイベント情報要約を活用する手法を提案する。Tsukada ら

[Prompt]

貴方に、{予測する日付}の日付と{会場名}という単語を含んだツイートの集合を与えます。この情報をもとに{予測する日付}に{会場名}で大規模なイベントがあるかどうか等を後述の形式で推測してください

注意点

- ・ イベントの有無の確信度は、0~100%の範囲で1%単位で答えてください
- ・ 別の日付や別の会場、ネット上のイベントのツイートがノイズとしてあります
- ・ ツイートから最も大規模と考えられるものだけ解答して、他の小規模イベントは無視してください
- ・ ツイート数の中央値は{会場ごとのツイート数中央値}個で、ツイート数が少ない場合はイベントがない可能性が高いです
- ・ 確信度が低い場合は不明と解答してください

解答の形式

イベントの有無の確信度：%
イベントの種類：
開始時間：hh:mm
イベント規模：12345人等
イベントの内容：内容の説明
理由：推論理由

以下からツイート（ツイート数{ツイートの数}）
{会場名と予測する日付を含んだツイートのリスト}

図 2 該当するツイート群を要約させるためのプロンプト。

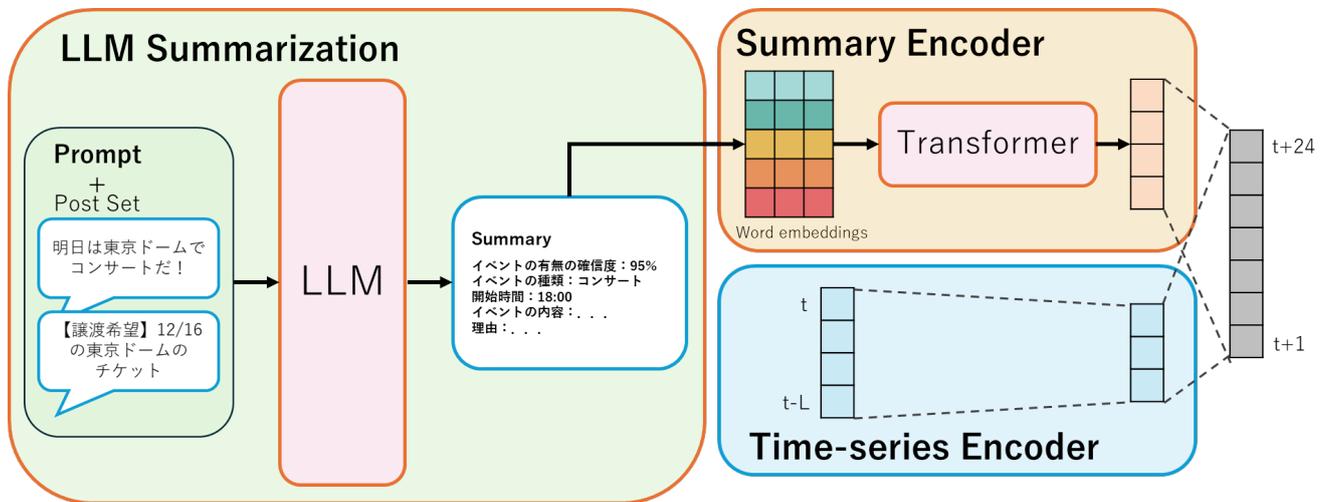


図 3 提案手法のモデルの概要。

は図 1 に示すように Transformer および Set Transformer を組み合わせた投稿エンコーダーと時系列エンコーダーからなるモデルを提案したが、本研究では、このモデルをベースに投稿エンコーダーを、LLM によるイベント情報要約モジュールと Transformer によるテキストエンコーダーに置き換えることで提案手法を実装する。

提案手法の説明のために、まず第 1 節でモデルのベースとして用いた Tsukadara らのモデルのアーキテクチャについて導入し、次に第 3.2 節で LLM 要約のために用いたプロンプトの解説を行う。そして最後に第 3.3 節で本研究の提案手法のアーキテクチャについて紹介する。

3.1 ベースライン手法

Tsukada ら [5] の提案した Set Transformer ベースのモデルを、本研究の提案手法のベース、および実験での比較手法として用いるため、まずそのマイクロブログ投稿の収集方法とモデルのアーキテクチャについて紹介する。

Tsukada らはイベント指標としてのマイクロブログ投稿の収

集の方法として、イベント会場 v において将来の日付 d の人口予測に用いる投稿は以下の条件をすべて満たすもの、という条件を設定した。

- ・ 予測日 d を表す表現を含む
- ・ イベント会場 v を表す表現を含む
- ・ 投稿が d よりも前に行われている。

それぞれを表す表現としては、 v ごとに表記ゆれを考慮した名寄せ辞書を作成し、日付表現に関しては「明日」のような投稿日と予測日の相対関係を考慮した検索クエリの拡張が行われた。

また、モデルのアーキテクチャとして、第 2.1 節で述べたように、Tsukada らはマイクロブログの投稿集合を入力として処理するために Set Transformer を組み込んだモデルを提案した。その概念図を図 1 に示す。Set Transformer は、Lee ら [7] が提案した集合の入力に対応し順序不変性を持つ Self Attention ベースのアーキテクチャであり、投稿間の複雑な関係性を考慮することが期待される。Tsukada らのモデルは、投稿と時系列それぞれのエンコーダー、および最終的な出力を得るための全結合層からなる。マイクロブログ投稿のエンコーダーでは、ま

表 1 LLM によるツイートからのイベント情報抽出の例

イベントの有無の確信度：95%
イベントの種類：プロ野球試合
開始時間：18:00
イベント規模：約 40,000 人（東京ドームの収容人数を考慮）
イベントの内容：読売ジャイアンツ対横浜 DeNA ベイスターズの試合。CS 進出に向けた重要な試合であり、「侍ジャパンナイター」として特別なイベントも行われる。
理由：多数のツイートが試合の詳細やチケット情報を含んでおり、特に「巨人対横浜 DeNA」という具体的な対戦が繰り返し言及されているため、確実に大規模なイベントが開催されると推測される。また、東京ドームの収容人数からも大規模な観客動員が期待できる。

ず各投稿は Word2vec モデル [8] により単語埋め込みが行われ、Transformer [9] により文章全体の埋め込みが生成される。その後、Set Transformer により各投稿の埋め込みから一つの埋め込みに変換される。時系列データのエンコーダーでは、イベント会場における短期的な人口変動のトレンドを捉えるために、1 時間を単位とした時系列 24 × 7 ステップ分が全結合層に入力される。最終的にこの 2 種のエンコーダーの出力が連結され全結合層により直後 24 ステップ（すなわち翌日の 24 時間分）の人口の時系列予測値が出力されるモデルとなっている。

本研究の提案手法では、このモデルの投稿エンコーダーを、LLM によるイベント要約モジュールとイベント要約のテキストエンコーダーに置き換えることでモデルを構築し、マイクロブログの投稿集合を時系列予測タスクに活用する。

3.2 LLM によるマイクロブログ投稿の要約

大規模言語モデルを用いてマイクロブログ投稿セットからイベント情報を抽出する手法を提案する。特定のイベント会場のある日付に関する多数の投稿全てを、プロンプトともに一度にまとめて LLM に入力し、その日にその場所で行われるイベントの情報を一定の形式を持った要約テキストとして出力させる。

第 1 節で述べたように、マイクロブログ投稿は文の形式が一定ではなく曖昧な表現で書かれるため、単独のツイートのみではその意図を正しく把握できないことも多い。また、第 3.1 節の方法で収集されたマイクロブログ投稿には、イベントとは無関係なノイズとなるような投稿も含まれており、予測精度の低下の一因となっている。

LLM は言語生成の能力により、広範なイベントの詳細を簡潔かつ適切に抽出することに優れている。また、LLM は Wikipedia をはじめとした膨大なインターネット上のデータをもとに事前学習されている [10] ため、アーティストの名前やイベントの名前のような特定の単語を識別し、事前学習により獲得した知識により補完する能力がある。そこで、図 2 に示すようなマイクロブログの投稿要約のプロンプトを設計する。このプロンプトではまず、LLM に与えられる情報の内容と、そこから予測日に特定のイベント会場で大規模イベントがあるかどうかを予測する指示が与えられる。また、さらに詳細の指示として、解答への注意点と解答形式が与えられ、最後に該当するマイクロブログ投稿とその数が続く。LLM のノイズ情報の除去の効率を高めるために、プロンプトには明示的にいくつかの

設定を追加した。まず、ノイズ投稿の存在を示唆したうえで、投稿内から最も大規模と考えられるイベントのみの情報を抽出し他のイベントの情報は無視するように指示を追加した。また、Tsukada らの研究 [4] において投稿数の多さが予測精度を改善する傾向が示唆されており、投稿数の多寡の判別の基準として各会場ごとの訓練期間における 1 日の投稿数の中央値を与えた。最後に、LLM に対して、投稿群から推測したイベントの有無においてより高い精度の情報を出力するため「ある」「なし」の 2 値分類ではなく「確信度」という百分率で出力させる指示を加えた。表 1 に実際の要約例を示す。

3.3 投稿要約を用いた人口変動の予測

LLM によるマイクロブログの投稿要約を用いて、イベント会場における人口変動を複数ステップにわたって長期予測する手法を提案する。図 3 にモデルの概要を載せる。提案するモデルは主に、LLM によるイベント要約モジュール、イベント要約のテキストエンコーダー、時系列エンコーダーからなる。入力として関連するマイクロブログ投稿の集合と直近の過去 24 × 7 時間分の人口時系列が与えられ、そのうち投稿群は前節で解説したプロンプトによりまとめて LLM に入力されイベント情報を抽出した要約として出力される。その後要約テキストは Transformer からなるテキストエンコーダーにより文章の埋め込みに変換される。このテキスト埋め込みと時系列エンコーダーによる時系列埋め込みを連結し、全結合層に入力することで直近将来 24 時間の人口時系列を予測する。

4 実験：LLM モデル毎の要約能力評価

提案手法で使用する LLM を決定するために、マイクロブログ投稿をもとにしたイベント情報の要約能力を複数の LLM において比較する予備実験を行った。比較として、LLM を用いたツイート要約単体におけるイベントの有無の予測精度をモデルごとに評価した。

4.1 実験設定

予測するイベント会場として東京ドームを設定し、2014 年 12 月から 2018 年 11 月までの約 4 年間を評価の対象とした。実験に使用したマイクロブログ投稿のデータは、研究室で収集している Twitter (現 X) データより、第 3.1 節と同様の条件と追加条件をもとに東京ドームという会場名を含みかつ予測する日付に対して言及している投稿を、会場名と日付の表記ゆれを考慮して収集した。追加条件は「予測日 d の前日の午前 0 時から午前 6 時の投稿の除外」であり、これはこの時間帯の「明日」等の日付表現を含む投稿が実際には予測日でなく前日を指す場合があり、それらの投稿がノイズとなる可能性が高いためである。最終的に収集したデータよりリポストなどによる重複を削除したデータを使用した。

次に各日付ごとのイベントの有無の正解ラベルの作成方法について説明する。作成には混雑統計データ¹と k -means 法を

1: 「混雑統計 ©」データは、NTT ドコモが提供するアプリケーション (※)

表 2 LLM によるツイート要約のイベント有無の精度

Model	Precision	Recall	F ₁
Llama-3-ELYZA-JP-8B	0.50	1.00	0.67
GPT-4o-mini	0.75	0.95	0.84

使用した。混雑統計データでは、250m 四方のメッシュとして地図を分割し、各メッシュにおいて単位時間内にその中に滞在した人数を混雑度と定義している。このデータには1時間単位で各メッシュの混雑度を推定した時系列データが含まれている。このうち東京ドームを含む特定のメッシュにおいて、2014年12月から2018年11月までの約4年間のデータを用い、最初の3年分を訓練データ、最後の1年分をテストデータとした。このデータを *k*-means 法でクラスタリングを行うことで正解ラベルの作成を行う。まず、各日付の時系列データを24次元の1つのベクトルとみなし、訓練データの各日付の時系列データのクラスタリングを行う。ほとんどの会場においてイベントがない日が多いことから、こうして形成されたクラスタのうち最も要素数が大きいクラスタに属する日付はイベントがない日、それ以外のクラスタに属する日付はイベントがある日と判定した。実際、非イベント日のクラスタは、平均ベクトルのピーク値が最も小さいクラスタであった。テストデータに対しては訓練データで作成したクラスタをもとに最近傍のクラスタを計算し正解ラベルの付与を行った。

事前学習済みの日本語 LLM モデルとして、OpenAI 社が提供する GPT-4o-mini モデル (OpenAI API より利用) [11] と、ELYZA 社が提供する Llama-3-ELYZA-JP-8B モデル [12] を比較する。これら2つのモデルに対して図2で示したプロンプトを用いてイベントの有無に対する確信度を百分率により回答させる形式でツイートの要約を行った。この確信度が一定の閾値を超えた日付に関してはイベントがあると判定しその分類精度を評価する。閾値の決定のために4年分のデータに対して前3年分のデータを訓練データとして最も F₁ score の高い閾値を計算し、最後の1年分のデータをテストデータとして2つのモデルの比較を行った。評価指標としては Precision, Recall, F₁ score を用いた。

4.2 結果

表2に LLM によるツイート要約のイベント有無の分類精度の結果をまとめた。Llama-3-ELYZA-JP-8B モデルは全てのテストデータにおいて同一の確信度を回答したため全てのテストデータにおいてイベントがあると分類され、分類器として機能しなかった。一方で、GPT-4o-mini モデルは F₁ score が 0.84 であり ELYZA のモデルよりも分類精度が高い結果となった。そのため、今後の実験におけるツイート要約の LLM として GPT-4o-mini モデルを使用する。また、特に prompt で指示していないが、イベントがある際に確信度を 100%側に推移

の利用者より、許諾を得た上で送信される携帯電話の位置情報を、NTT ドコモが総体的かつ統計的に加工を行ったデータ。位置情報は最短5分毎に測定される GPS データ (緯度経度情報) であり、個人を特定する情報は含まれない。※ドコモ地図ナビサービス (地図アプリ・ご当地ガイド) 等の一部のアプリ。

表 3 対象とした8イベント会場。各会場の種別、対象期間4年間のイベント日数、訓練期間のツイート数の中央値を示す。

種別	会場名	イベント日数	ツイート数中央値
サッカー場	日産スタジアム	79	0
	味の素スタジアム	132	2
野球場	東京ドーム	712	39
	明治神宮野球場	538	2
	横浜スタジアム	343	5
展示会場	パシフィコ横浜	339	5
コンサート会場	横浜アリーナ	557	12
	日本武道館	497	13

させ、ない際に0%側に推移させることが確認できた。

5 実験：人口変動の時系列予測と分類評価

提案手法の人口時系列予測における時系列予測誤差と、時系列出力をもとにしたイベントの有無の分類の評価を行う。

5.1 実験設定

対象とするイベント会場として、関東圏から8会場を選出した。各会場の情報を表3に示す。実験で使用したデータセットは、イベント会場における人口統計データとマイクロログ投稿データの2つである。実験では、イベント会場における人口の時系列データとして、第4節の予備実験に引き続き混雑統計データを使用した。各会場を含む特定のメッシュのデータに対して、2014年12月から2018年11月までの約4年間のデータを用い、最初の2年分を train データ、次の1年分を dev データ、最後の1年分を test データとして使用した。また、マイクロログのデータとしても同様に予備実験で用いた条件をもとにデータを収集した。これらのデータをもとに [5] の設定に従って train データのオーバーサンプリングを行い合計200エポックのデータで学習を行った。

評価において、ベースラインとして Tsukada ら [5] が用いた Set Transformer (以下 ST) ベースのモデルを用い、提案手法との比較を行った。またイベントの有無の評価においては、第4節で行った LLM の要約モジュール単体による分類評価も追加で行った。提案手法において LLM 出力のイベント要約は日本語形態素解析器 JUMAN++ [13] によって単語を分かち書きしたのち、主に SNS と wikipedia データによって学習された hottoSNS-w2v [8] によって単語埋め込みに変換され Transformer に入力することで要約の文埋め込みを得る。

5.2 評価指標

人口の時系列予測の評価として、まず実際の人口変動をどの程度正確に予測できているかを計るため式1で表される加重絶対パーセント誤差 (WAPE) を用いて評価を行った。

$$WAPE = \frac{1}{D} \sum_{d=1}^D \left(\frac{1}{24} \sum_{t=1}^{24} \frac{|\hat{X}_{dt} - X_{dt}|}{\bar{X}_{dt}} \right) \quad (1)$$

ここで、 X_{dt} , \hat{X}_{dt} はそれぞれ日付 d の時刻 t における実測値と

表 4 人口時系列予測での WAPE 評価

会場	WAPE(test 期間)		WAPE(イベント日)	
	ST モデル	提案手法	ST	提案手法
日産スタジアム	0.800	0.712	0.521	0.509
味の素スタジアム	0.833	0.797	0.662	0.697
東京ドーム	0.384	0.415	0.338	0.309
明治神宮野球場	0.629	0.585	0.503	0.480
横浜スタジアム	0.371	0.387	0.414	0.411
パシフィコ横浜	0.437	0.456	0.403	0.397
横浜アリーナ	0.370	0.382	0.451	0.405
日本武道館	0.720	0.674	0.670	0.610
平均	0.568	0.551	0.495	0.477

表 5 人口時系列予測でのイベントの有無の精度評価

会場	F ₁ score		
	ST モデル	LLM 要約単体	提案手法
日産スタジアム	0.667	0.652	0.800
味の素スタジアム	0.567	0.623	0.625
東京ドーム	0.849	0.851	0.845
明治神宮野球場	0.763	0.664	0.776
横浜スタジアム	0.767	0.800	0.856
パシフィコ横浜	0.465	0.379	0.553
横浜アリーナ	0.469	0.544	0.510
日本武道館	0.385	0.533	0.476
平均	0.617	0.631	0.680

予測値であり、 \bar{X}_{dt} は実測値の平均、 D は評価対象の日数である。本研究はイベント会場における人口予測が目的であり、イベントの日数は会場や評価期間により変動してしまい全テスト期間のみの評価では精度を過小評価する可能性がある。そのため WAPE の評価は全テスト期間、およびテスト期間中のうち実際にイベントがある期間の 2 つの評価を行う。

また、イベント会場での人口予測においてはイベントの有無の予測も重要であるため、時系列出力をもとに各予測日のイベントの有無を分類し F₁ score によって評価を行う。時系列出力からイベントの有無を分類する手法として、第 4 節で用いた k -means 法をもとにした手法を使用する。正解ラベルの作成時と同様に、train 期間と dev 期間の実際の時系列データをもとに k -means 法でクラスタリングを行う。こうして形成されたクラスタを用いて、提案手法のモデルが出力した 24 時間分の予測時系列に対してイベントの有無の分類を行う。

5.3 結果

最初に、人口変動の時系列予測の WAPE 評価の結果を表 4 に示す。test 期間全体に対する予測誤差評価においては 8 会場中 4 会場において提案手法が上回っているが全体において ST 手法と同程度である。一方、イベント日のみの誤差評価においては提案手法が 7 会場において ST 手法を上回り、実際のイベント発生時の予測精度が既存手法より高いことが示された。

また、イベントの有無における 2 値分類としての精度評価

の結果を表 5 に示す。LLM 要約モジュール単体の分類精度と ST 手法の精度を比較すると、平均で 1.4% 精度が向上しており、LLM による要約のみでも ST モデルと同等の分類能力があり、イベント情報抽出における LLM の高い文章生成能力の有用性が示された。さらに、ST 手法と提案手法を比較すると、8 会場中 7 会場で提案手法が ST 手法を上回り、平均で 6.3% 精度が改善した。これより LLM によるノイズ情報除去やイベント情報の抽出の有用性が示された。また、LLM 要約単体に対しても 4.9% 精度が向上しており、要約のイベントの有無の確信度による分類が間違えた場合であっても、LLM を追加学習させずに外部の比較的小規模なモデルにイベントの種類やイベントの内容といった要約情報を学習させることで精度向上が可能であることが示された。

会場別に結果を見る。最も予測誤差が小さい東京ドームにおいては他の会場と異なり ST 手法に対する予測誤差や分類精度の改善が見られない。東京ドームはツイート数もイベント日数も最も多い会場であり、ST 手法の段階で高い分類精度が達成されており、マイクロブログ投稿による予測精度改善による限界の可能性が示唆される。また、ST 手法に対する LLM 要約単体での F₁ score の向上が最も高いのは、横浜アリーナと日本武道館の 2 つのコンサート会場である。コンサート会場は、Tsukada らによりアーティストなどの固有名詞が多いわりに多様なアーティストに利用されるため各単語の登場頻度が低く予測精度が低い点が指摘されていたが、これらに対して LLM による情報の補完と文章のフォーマット化が有効であることが示された。一方でコンサート会場は他の会場と異なり提案手法の F₁ score が LLM 要約単体と比べて大きく減少しており、ST モデルと同様に低頻度の単語による精度低下の可能性が示唆された。

6 まとめ

本研究では、マイクロブログ投稿を用いたイベント会場周辺における人口変動の予測タスクにおいて、投稿群からのイベント情報抽出とノイズ除去のために LLM を活用する手法を提案した。提案手法は LLM を用いることで投稿群からイベント情報を要約テキストという形で得ることで結果の解釈性が向上しており、また時系列予測の手法としては多くの会場において従来手法と同等以上の精度を達成した。また、多数の投稿を同時に処理する ST 手法やパラメータ数の多い LLM の追加学習と異なり、提案手法は学習コストが低い。今後の研究方針は、使用する LLM モデルによる精度への影響や、1 週間単位の長期間イベントにおいてツイート数が少なくなる中間日などの精度の改善などが考えられる。

7 謝辞

本研究成果は、国立研究開発法人情報通信研究機構 (NICT) の委託研究 (222C02) および、国立研究開発法人情報通信研究機構 (NICT) の委託研究 (JPJ012368C05401) により得られたものです。

文 献

- [1] 塚田涼太郎, 詹浩森, 石渡祥之佑, 豊田正史. マイクロブログおよび携帯電話人口統計を用いた大規模イベント会場における人口変化の長期予測. 第12回データ工学と情報マネジメントに関するフォーラム, 2020.
- [2] Binbing Liao, Jingqing Zhang, Chao Wu, Douglas McIlwraith, Tong Chen, Shengwen Yang, Yike Guo, and Fei Wu. Deep sequence learning with auxiliary information for traffic prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, p. 537–546, New York, NY, USA, 2018. Association for Computing Machinery.
- [3] Tatsuya Konishi, Mikiya Maruyama, Kota Tsubouchi, and Masamichi Shimosaka. Cityprophet: city-scale irregularity prediction using transit app logs. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, p. 752–757, New York, NY, USA, 2016. Association for Computing Machinery.
- [4] 塚田涼太郎, 詹浩森, 石渡祥之佑, 豊田正史, 梅本和俊, 商海川, 是津耕司. 未来のイベントに言及するマイクロブログ投稿を用いた人口変化の予測. 第13回データ工学と情報マネジメントに関するフォーラム, 2021.
- [5] 塚田涼太郎, 豊田正史, 梅本和俊, 是津耕司. イベント発生の不均衡性に適合した人口変動予測. 第16回データ工学と情報マネジメントに関するフォーラム (DEIM 2024), 2024.
- [6] Yuebing Liang, Yichao Liu, Xiaohan Wang, and Zhan Zhao. Exploring large language models for human mobility prediction under public events. *Computers, Environment and Urban Systems*, Vol. 112, p. 102153, 2024.
- [7] Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, Vol. 97 of *Proceedings of Machine Learning Research*, pp. 3744–3753. PMLR, 09–15 Jun 2019.
- [8] 松野省吾, 水木栄, 榑剛史. 日本語大規模 sns+web コーパスによる単語分散表現のモデル構築. 人工知能学会全国大会論文集, Vol. JSAI2019, pp. 4Rin113–4Rin113, 2019.
- [9] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.
- [10] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 2023.
- [11] OpenAI. Gpt-4o-mini. <https://platform.openai.com/docs/models/gpt-4o-mini>, 2024. Large language model.
- [12] Masato Hirakawa, Shintaro Horie, Tomoaki Nakamura, Daisuke Oba, Sam Passaglia, and Akira Sasaki. elyza/llama-3-elyza-jp-8b, 2024.
- [13] Arseny Tolmachev, Daisuke Kawahara, and Sadao Kurohashi. Juman++: A morphological analysis toolkit for scriptio continua. In Eduardo Blanco and Wei Lu, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 54–59, Brussels, Belgium, November 2018. Association for Computational Linguistics.