

平成 27 年度 修士学位論文

題目 途中立ち寄る事を前提とした
経路を取得する事に関する研究

指導教員 服部 峻

提出者 室蘭工業大学大学院 情報電子工学系専攻

氏名 永澤 勇樹

学籍番号 14043031

提出年月日 平成 28 年 1 月 29 日

目次

第1章	序論	1
第2章	関連研究及び既存のシステム	3
2.1	経路検索	3
2.2	混雑度測定	4
第3章	システムの提案	6
3.1	システムの概要	6
3.2	使用例	7
3.3	途中地点とはどのようなものか	7
3.4	途中地点を選ぶ基準	8
3.5	提案システムの詳細	9
3.6	途中に立ち寄る場所と混雑度の関係	12
第4章	SNS を用いる有用性	13
4.1	ツイートから情報を抽出	13
4.2	位置情報による検索	14
4.3	今居るツイートの調査結果	16
第5章	混雑度の調査	19
第6章	混雑度の予測	21
6.1	ツイート数の取得方法	21
6.2	ツイート数の結果と考察	24
6.3	ツイート数をそのまま用いた単回帰分析	25
6.4	残存効果の適用	28
6.5	時間帯を分けて残存効果を適用	32
第7章	結論	35
	謝辞	37

目次

1.1	途中に立ち寄る事を考慮した経路検索のシステム構成と本研究で取り上げる範囲	2
3.1	途中に立ち寄る場所から経路を決める例	6
3.2	途中に立ち寄る事を考慮した経路検索システムの使用例	10
4.1	1月12日に札幌駅から半径0.5km以内で投稿されたツイート	14
4.2	駅にいるツイートから飲食店へ向かっている例	15
5.1	6月10日の店内人数とツイート数	20
5.2	6月12日の店内人数とツイート数	20
5.3	6月14日の店内人数とツイート数	20
5.4	6月15日の店内人数とツイート数	20
5.5	7月29日の店内人数とツイート数	20
6.1	Swarmから投稿されたツイートの例	23
6.2	文章で今居る事が読み取れるツイートの例	23
6.3	写真が載っている今居る事が読み取れるツイートの例	23
6.4	会話の内容から今居る事が読み取れるツイートの例	23
6.5	店名を用いないで今居る事が読み取れるツイートの例	23
6.6	写真のみの今居る事が読み取れるツイートの例	23
6.7	6月10日の単回帰分析に基づく店内人数と実際の店内人数	26
6.8	6月12日の単回帰分析に基づく店内人数と実際の店内人数	26
6.9	6月14日の単回帰分析に基づく店内人数と実際の店内人数	27
6.10	6月15日の単回帰分析に基づく店内人数と実際の店内人数	27
6.11	7月29日の単回帰分析に基づく店内人数と実際の店内人数	27
6.12	6月10日の残存効果適用による予測店内人数	29
6.13	6月12日の残存効果適用による予測店内人数	29
6.14	6月14日の残存効果適用による予測店内人数	30
6.15	6月15日の残存効果適用による予測店内人数	30

6.16	7月29日の残存効果適用による予測店内人数	30
6.17	6月10日の時間帯を分けた予測店内人数	33
6.18	6月12日の時間帯を分けた予測店内人数	33
6.19	6月14日の時間帯を分けた予測店内人数	33
6.20	6月15日の時間帯を分けた予測店内人数	34
6.21	7月29日の時間帯を分けた予測店内人数	34

表目次

4.1	札幌駅周辺で発信されたツイートの調査結果	17
6.1	ツイート数と店内人数及び占有数の相関係数	25
6.2	前の時間のツイート数の効果を適用したツイート数の相関係数	28

第 1 章

序論

経路を決定する時には、様々な要素が決定要因となる。例えば、目的地まで早く着く、乗り換えが少ない、費用が掛からない等である。また、到着時刻と所要時間も要素となるであろう。しかし、YAHOO!路線 [1] やジョルダン [2] などの多くの経路検索エンジンでは通る経路または乗車する交通機関に関する情報のみ書かれている事が多い。移動中に見られる風景や、乗り換え時間中にできる事は重要視されていない。

途中に通る場所でも絶景等、有意義な情報がある場合もあるが、従来の経路検索エンジンでは殆ど取り扱わない。絶景スポット等を予め知っているならば優先して通る経路検索結果を選ぶ、またはそこを通るような結果を出すように検索条件を入力できるが、知らない場合は知る機会が無いので見向きもされない事があり得る。また、立ち寄るまたは通る場所を選ぶ必要があるのではないかと考えた。食事で例えると、寄る事ができる飲食店を選んで、それを元に経路を決めるものである。または移動中に桜並木等絶景が見られる所を通り、移動しながら楽しむ経路を選択するというのもあり得る。以上の事から途中で体験できる事、例えば途中で下車して食事する事や途中で見られる風景なども経路を決定する要素と成り得ると考えられる。

途中に立ち寄る場所を決定する判断基準は多数ある。例えば食事のみで考えた場合、立ち寄る時点における時刻、最寄り駅から立ち寄る場所までの距離と所要時間、天候、食べられる食事、値段、評判、混雑などが挙げられる。以上の多数の判断基準のうち、本研究では特に「混雑」に注目した。混雑によって、想定した時間以上に時間が掛かり予定が狂う事、また体験を断念する事があるからである。また、旅行等で寄る場所を決める際に、リアルタイムで得た混雑度が Web サイト等に載っている事が殆ど無いというのも、混雑に注目した理由である。混雑を測定する既存手法はいくつかある。カメラ等ハードウェアを用いて、混雑を測定したい場所をその場で物理的に観測する手法である。この方法では、店の許可または事前準備が必要であり、また、来客の許可を得る必要もある。そのため、ツイッター [3] 等 Web 上で公開されている情報を用いて現実世界での物理現象、特に混雑度を予測する手法を検討した [4, 5]。具体的には投稿者が自身の居場所を知らせている事を読み取れるツイートの数を用いて単回帰分析を行い、その場所の混雑度を予測するという手法を試みた。

本研究では途中に立ち寄る場所を考慮した経路検索に必要な仕様を述べて、ユーザが途中地点を判断するための要素が多数ある中、混雑度を取り上げた。SNS を用いて混雑度を取得するために、今居る事を読み取れるツイートの総数を用いる事を試みた。途中に立ち寄る事を考慮した経路検索のシステム構成と本研究で取り上げる範囲を図 1.1 に示す。赤い実線で囲んだ部分が本研究で取り上げる範囲である。

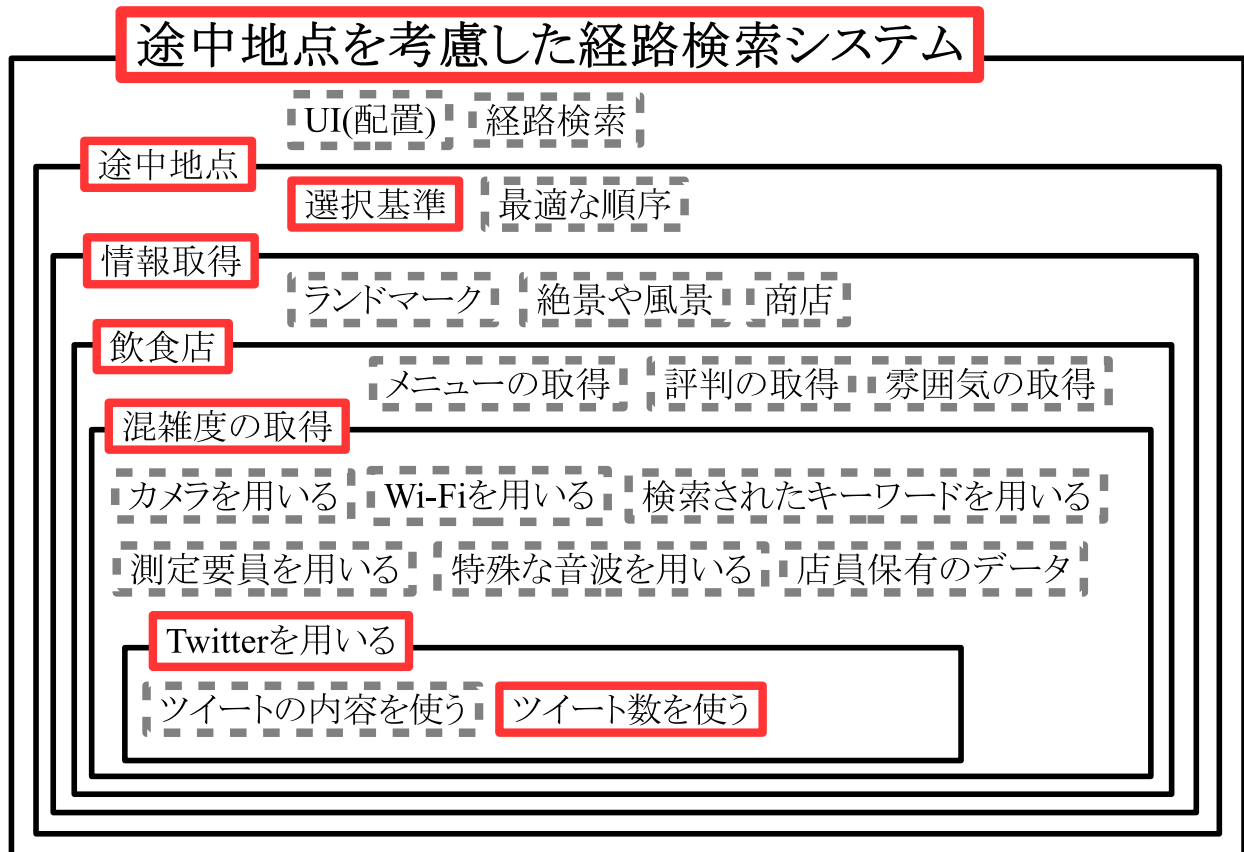


図 1.1 途中に立ち寄る事を考慮した経路検索のシステム構成と本研究で取り上げる範囲

本研究で行った事をまとめると、多くの経路検索エンジンではあまり考慮していない途中地点の情報を取得する事を行った。システムを構築するにあたって、途中地点をユーザに推薦する必要があると考え、情報取得を主に行った。数多くの途中地点のうち、最も利用が多いであろう飲食店に注目した。飲食店を推薦する要素として、本研究では混雑度に注目して、取得を行った。混雑度の取得として多数の手法がある中、ツイッターのツイート数を用いて混雑度の取得を行った。

第 2 章

関連研究及び既存のシステム

本章では，経路検索と混雑度測定に関する従来手法の問題点について挙げ，提案手法の有用性についても述べる．

2.1 経路検索

多くの経路検索エンジンは，途中に寄る場所を考慮していない事が多い．例えば，札幌～名寄をジョルダンの「18 きっぷ検索 [2]」を用いて経路検索した場合，旭川で 135 分の乗り換え時間が発生する．乗り換え時間がある事以外に情報は殆ど無く，乗り換え時間中の過ごし方については考慮されていない．そのため，135 分の待ち時間を過ごす方法は自力で調べる，あるいは，旅行記等を検索してユーザが自分で調べる前提だと思われる [6]．

また，従来のシステムでは乗り換え場所または立ち寄り場所毎に候補を選択する事ができない．経路の候補が複数得られるものの，これらは複数の経路から選ぶための候補であり，乗り換えまたは立ち寄り場所を元を選ぶわけではない．移動の途中に寄る事または発生する出来事，例えば食事や休憩の時間が特定の時間に取れる場所，移動中に見られる景色，季節限定のイベントや催し物が行われている場所等を予め知る事で，ただ移動するだけの旅を楽しむ事ができたり，安心した旅にする事ができる．また，逆に知らなかった事で休憩できる場所が無い状態が長く続く，何も無い場所で長時間待たされる，狭部を通る，食事を取れない等，厳しい旅になる可能性もある．しかし従来のシステムでは，経路上に何があるか，またはどこを通るかを意識して選ぶ事は殆ど無いだろう．

それでも予め途中に何があるか知っていれば途中の場所を選択して道中を楽しむ，または苦難を避ける事ができるが，初めて通るあるいは月日が経った等，途中に何があるか不明な状態では，選ぶのは困難だと考えられる．途中の経路を調べる場合でも経路毎に，通る場所について寄る価値のある場所があるかどうか調べる必要がある．また通って眺めるならば問題無いが，食事や休憩のように時間が掛かる場合があり，寄り場所を決めた後に，途中地点で掛ける時間を考慮して経路を再検索する必要があり，二度手間となる．

本研究で提案する途中地点を考慮した経路検索システムによって，何処を通るか，危険な箇所や無補給区間がある事を予め知る事でそれらを回避し，移動における苦難を避ける事ができ

ると考えた。また、通る場所で行われているイベントや見られるもので選ぶ事によって、目的地以外で楽しむ事ができ、移動中を含む旅行全体をより楽しむ事ができるだろうと考えた。それによって、地域における新たな発見ができるのではないかと考えた。

2.2 混雑度測定

本研究では途中地点を考慮した経路検索において、途中地点をユーザが判断する要素として混雑度の取得を行った。混雑度を測定する手法として以下が考えられる。

1. Wi-Fi 等、端末から発信される電波から店内の端末数（客数）を推測 [7, 8]
2. 特殊な音波を発信して端末のカメラで集音する [9]
3. 店内カメラの映像から推測 [10, 11]
4. 測定要員を配置
5. SNS にある今居る事が分かる文章から推測

1 と 2 はハードウェアを用いて各個人が持っている端末の数を測る手法である。総務省のデータ [12] によると、スマートフォンの世帯普及率は 49.5% であり、今後も増えていく事から端末の数を数える手法は有効であると考えられる。公衆 Wi-Fi を用いて MAC アドレスを収集して端末を追跡するターゲティング広告が行われた例がある [13] 事から、特定のアクセスポイントに接続した端末数を数える事で、混雑度を測定できるのではないかと考えた。しかし、電波が広範囲に広がる事から、密室や広域における測定では有効だと考えられるが、狭い区域にアクセスポイントが密集している状況においては、他の場所の電波を拾う事や混信による誤認識が多くなると考えられる。音波を使って混雑度を測る手法は、JR 東日本アプリ [14] の山手線トレインネットの混雑度状況で用いられている。車両毎に異なる音波を用いて、各端末のマイクで集音した数を数える事によって、各号車の混雑率が分かる仕組みである。鉄道車両のように仕切られた環境においては、隣の車両から発する音を拾う事は無いが、開放された空間においては誤認識の可能性があると考えられる。

3 と 4 はハードウェアを用いて人の流れを測定する手法である。カメラを用いた手法については、カメラに映っている人をシステムが判断する事で混雑度測定が行えるだろう。カメラを用いた人の判別については、事前に怪しい動きを感知して万引きや暴力等の犯罪を防ぐ方法が研究されている [15]。混雑度の測定においては、店の出入り口付近の人の流れまたは店内にいる静止した人を判断できれば十分である事から、容易に行えると考えられる。

1 から 4 はハードウェアを用いる手法であり、実績があり信頼度は高い。しかし、これらの手法で測定するには必要な機器を用意しなければならないため、導入していない場所においては測定ができない。また、この手法ではいずれも店側が一方的に客からデータを収集して公開する形となる。プライバシーの問題に関わるため、事前に客への説明または同意が必要となるだろう。また混んでいる情報が提示される事によって、空いている他の店に客が流れる可能性があり、自店の混雑度を公開するのはあまり好ましくないと考える店もあるであろう。

5 つ目の手法は SNS に各ユーザが投稿し公開された内容を用いて混雑度を測る手法である。本研究で想定しているのはツイッターで、自分の居場所を自ら投稿する事がある。また、Swarm[16] 等、位置情報の発信を行う外部アプリがあり、容易に位置情報の発信を行う事ができることから、位置情報を発信するユーザは多いと考えられる。位置情報は GPS による位置情報の他、その場所で見られる風景、場所や住所名を書く等によって表されている。そのため、リアルタイムでそのユーザが今どの辺りに居るかが分かるのである。人気の場所においては同じ場所に居るツイートが複数のユーザから発信される事がある。これを利用して、自宅が留守である事や外出中である事を察知されて空き巣に狙われやすい事が問題となったが、同じ場所から発信しているツイートの多さにより、混雑している事が分かると考えた。また、自ら情報を発信している事から、プライバシーの問題にはならないであろう。自らここに居る事を発信しているツイートを用いて、ある施設における混雑度を測れないかと考えた。

ツイッターは非公開アカウント以外のツイートを検索する事ができる。施設名を用いて検索した結果から、今居る事が分かるツイートのみを抜き出して混雑度を測る。予想としては飲食店の場合、昼食時と夕食時に混む事が多いため、今居る事を示すツイートの数もそれと相関があると考えた。

第3章

システムの提案

本章では、途中に立ち寄る事を考慮した経路検索システムの提案について述べる。

3.1 システムの概要

本研究で提案するシステムは、目的地まで行く経路を、目的地に着くまでの途中の地点での出来事や施設から経路を決めるシステムである。例えば出発地と目的地を入力したら、図 3.1 のように出発地から目的地までの経路中で、途中に立ち寄れる場所の候補が 5 つ表示される。一つの経路当り途中に立ち寄る場所が一つをすると、選べる経路は 5 通りである。5 つのうち 1 つの立ち寄る場所の候補を選択すると、選んだ場所を通りかつそこを楽しむのに十分な時間が与えられた経路が出力される。このシステムは、主に目的地が予め決められており、多少遅く着いても問題が無い場合を想定している。また、目的地へ向かう途中で、途中の場所に立ち寄ってその場所で行われているイベントまたは見られる景色を楽しむ事を目的としている。

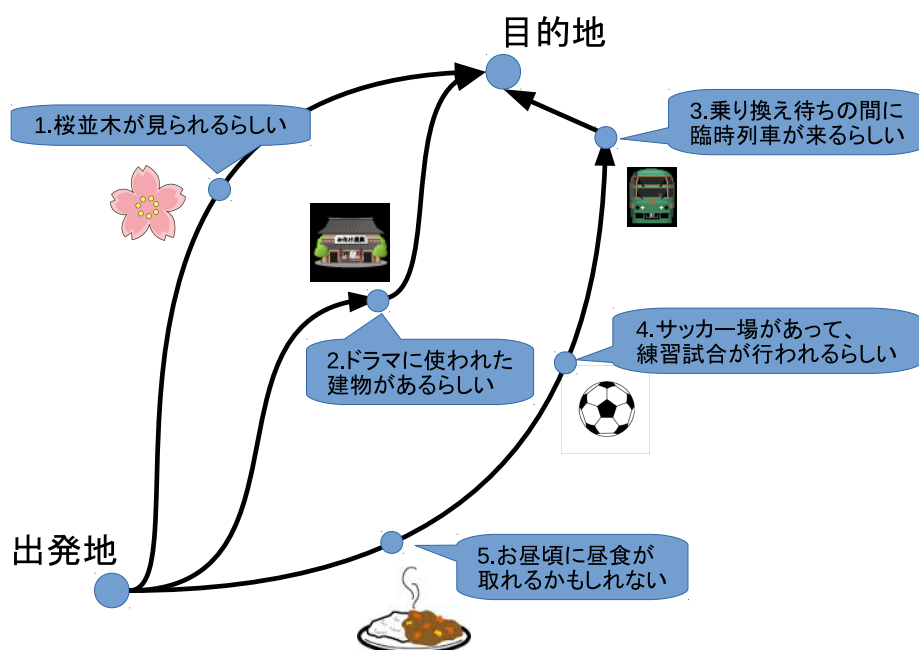


図 3.1 途中に立ち寄る場所から経路を決める例

3.2 使用例

本節では、提案するシステムの使用例を挙げる。

例えば、帰省で室蘭工業大学近くの自宅から札幌の実家まで行く場合を想定する。当初、途中どこにも寄らないで札幌の実家に行く事を考えた。しかし、到着に余裕があり遅い時間に着いても構わないので、道中で何か面白い事が無いかと思っていた。このシステムを用いて道中にイベントか何かが無いか探したところ、途中の苫小牧にて本日限りで今大人気のゲーム大会がある事を知り、参加したいと思った。システムによって、苫小牧に寄ってゲーム大会を楽しめる滞在時間を含んだ経路検索結果を元に札幌の実家へ帰省することにした。帰省当日、途中の苫小牧で途中下車して参加して、今日しかできない貴重な体験をした。

もう一つの使用例を挙げる。オープンキャンパスで室蘭工業大学及び室蘭の観光に初めて行く状況で、室蘭のみならず途中に通る場所についても知らない状態を想定する。10時に札幌のホテルを出て、15時に室蘭工業大学に到着する見込みだと仮定する。12時を超えるので途中で昼食を取る計画であるが、初めて通る事もあり、12時頃に何処を通りどのような食事が取れるかは不明である。このシステムを用いて、検索条件として12時頃に昼飯を取りたい事を入力することで、12時台に何処で何が食べられるかを元に経路を決めて、室蘭へ行く事にした。

その他に、列車に乗っている途中で車窓から桜並木または紅葉を見られる経路、また特定の列車とすれ違う等移動中に特定の物を見たい時にも応用ができると考えられる。

以上の例をまとめると

- 途中どこかに寄れる余裕がある時
- 途中どこかで何かしたい要望がある時
- 移動中に見たいものがある時

上記の3つのうち、どれかを満たす時に有用性があるのではないかと考えている。途中に通る場所を考慮していない従来の経路検索システムでは、道中の面白いイベントや途中で食事を行う機会が無く目的地へ行ってしまふかもしれない。

3.3 途中地点とはどのようなものか

ここで言う「途中地点」とは、旅行経路の出発地点から目的地の間にある点である。ただ観る、体験する、消費する等、当初から訪れたかった場所ではないが楽しめる場所である。また、どこかに降りて滞在するものではなく、移動中に車窓から見られるというものもある。途中地点の例としては「温泉」「食事」「宿」「特定の列車」「絶景」等がある。

本来行く場所ではない事もあり、その場所に関わる時間は目的地に滞在する時間よりも少ない。そのため、滞在が少ないのに膨大な時間を掛けて調べる事は、旅行計画として非効率となるので、あまり調べていない事がある。

3.4 途中地点を選ぶ基準

途中に立ち寄る場所の候補は多数ある。特に絶景や風景等、基準が曖昧または人によって基準が異なる場合は、全ての場所が立ち寄る場所の候補と成り得る。そのため、候補として出力する立ち寄る場所を絞る必要がある。例えば、どのような目的で立ち寄るかを経路検索を行う前に入力する等の対策が必要である。全ての場所が立ち寄るのに最適な場所と成り得るので、絞る必要がある。

立ち寄る場所毎に適切な滞在時間は異なる。ジャンルごとに異なると言っても良いだろう。例えば、モニュメントを眺めるのに必要な時間と食事を取るのに必要な時間は異なる。滞在時間が短いと、立ち寄る場所で体験中に出発時刻となってしまう、途中で放棄または出発時間を延ばす事になりかねない。また、無駄な時間が発生するため、極端に滞在時間が長いのも良くない。

立ち寄る場所を選ぶ基準として以下が考えられ、またその他に様々な要素が考えられる。

- ユーザの嗜好
- イベントの有無
- 時刻
- 天候
- 場所

例えば、イベント公演がある所に行く、12時頃に着く場合は食事を取れる場所に行く、悪天候の場合は屋外に出ない、駅からの距離等である。また、一度行ったことがあるか否かというのも選ぶ基準になると考えられる。限定販売や本日限りのイベント等希少性や、話題になっているかどうか等話題性も関係があると考えられる。

ジャンルごとに異なる別の選択基準がある。例えば飲食店の場合、立ち寄る店を選択するのに必要な情報として以下のようなものが考えられる。

- 外観
- 食品とメニュー
- 値段
- 雰囲気
- 評判
- 混み具合

これらの選択基準の多くは公式サイトから得られ、また多くの既存データベースからも得ることができる。一方で一部の選択基準はリアルタイムで変化しており、立ち寄った時刻における情報を得る必要がある。

3.5 提案システムの詳細

この項目では、途中に立ち寄る事を考慮した経路検索システムの構成案について述べる。

3.5.1 システム構成案

本章では、立ち寄る場所の候補を大幅に絞るため鉄道旅行で、飲食店に立ち寄る前提で考える。最初は、途中に立ち寄る場所を決めるために必要な以下の項目を入力する。

- 出発する場所
- 目的地
- どのような場所に立ち寄りたいか（立ち寄る目的）

また、オプションとして「出発時刻」「到着時刻」「立ち寄る時刻」「要素」「持ち時間の余裕」も考えられる。「立ち寄る目的」は「温泉」「食事」「宿」「特定の列車」「絶景」等であり、数多い立ち寄る場所の候補を絞る。「要素」は例えば食事と言えば「カレー」「ラーメン」等ジャンルであり、選んだ「立ち寄る目的」によって選択肢が変わる。「持ち時間の余裕」は立ち寄る場所における時間の余裕である。時間の余裕をどれほど持つかを選択する項目で、3段階で設定する事が良いと考えている。この項目は例えば食事等、個人によって滞在に必要な時間が異なる場合があり、調整用の項目を入力の段階で調整する必要があると考えた。

オプションの「出発時刻」「到着時刻」「立ち寄る時刻」は、全てを入力する必要は無い。候補から選んだ途中に立ち寄る場所に着く時刻から決まる事もあるからである。これらを入力する主な理由は、同じ場所でも時刻によって最適かどうかが変わる事があるからである。例えば、明るい時間に絶景が見られる場所でも、日没後の暗い時には殆ど見えない事があり得る。その他、特に店が関わる場合は閉店時に寄る事を避けるために、営業時間を考慮に入れる必要がある。

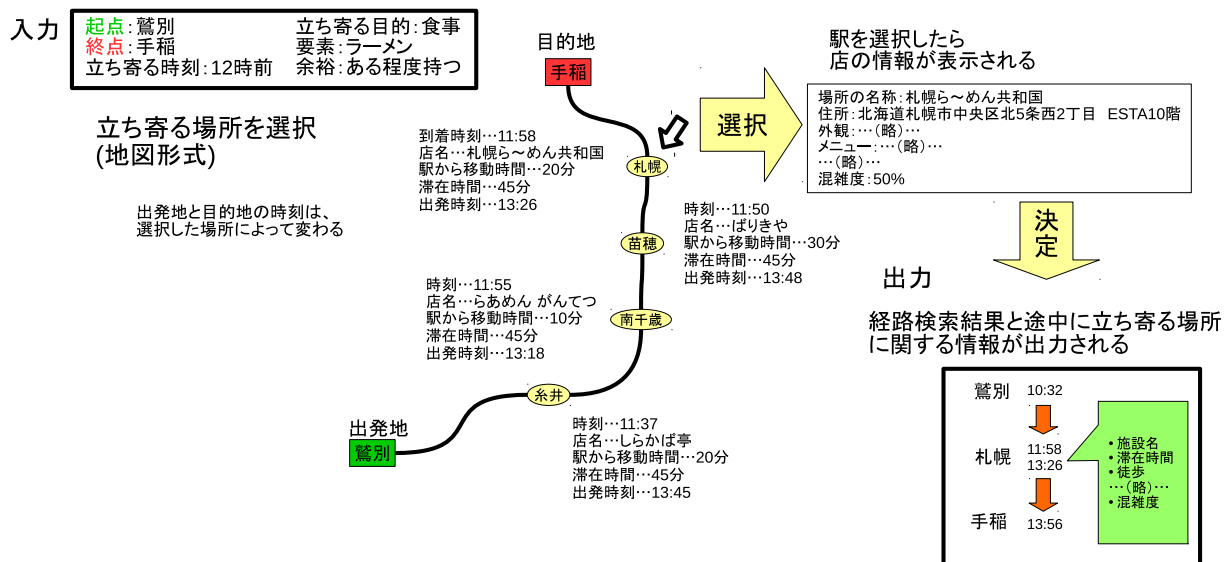


図 3.2 途中で立ち寄る事を考慮した経路検索システムの使用例

上の図 3.2 は、提案したシステムの図解である。鷺別から手稲まで行くと仮定し、途中 12 時前にラーメンが食べられる店に寄る設定である。「出発時刻」と「到着時刻」は設定していないが、選択した途中で立ち寄る場所によって両時刻が決まるとする。条件を入力した後は、最初に入力した「立ち寄る目的」に適した「途中で立ち寄る場所」の候補が複数表示される。場所の候補は地図上でプロット、あるいはリスト形式で表示される。図 3.2 の場合は、系井駅の「しらかば亭」、南千歳駅の「らあめん がんばてつ」、苗穂駅の「ばりきや」、札幌駅の「ら〜めん共和国」が途中でラーメンを食べられる場所として表示されている。

最初に表示される情報は「場所の名称」「その地点までの所要時間」「滞在時間」「出発時間」が表示され、ユーザはこの情報を元に複数の候補から選ぶ。図 3.2 の場合は、候補を選択した後は詳細情報がリストで表示されて、この場所で本当に良いかどうかを決定させる。例えば、札幌駅（ら〜めん共和国）を選択すると、名称、住所、メニュー、到着時の推定混雑度等を表示して、ここに寄るかどうかを決めさせる。詳細情報は選んだ「立ち寄る目的」によって異なる。例えば飲食店の場合、「外観の写真」「食品とメニュー」「値段」等である。選んだ後は出発地からその場所に寄り、また楽しむのに十分な時間を持って目的地へ向かう経路検索結果が出力されるシステムである。図 3.2 の場合は、鷺別駅を 10 時 32 分に発車して札幌駅には 11 時 58 分に到着する。その次は「ら〜めん共和国」で食事をする時間である 85 分を確保する。その後、札幌駅を 13 時 26 分に発車して目的地の手稲駅に 13 時 56 分に到着する経路が出力される。

3.5.2 システムの挙動

本項では立ち寄る場所を見つける手法について説明する。手順は以下の通りである。

1. 起点から終点までの経路を検索
2. その経路の途中にある駅周辺にある施設を検索

3. 見つかった施設全てについて詳細情報を検索
4. 駅の数だけ 2,3 を繰り返す
5. 入力した情報を元に立ち寄る場所の候補をユーザに提示する

必要事項及びオプションを入力した後は、「出発する場所」から「目的地」への何処にも寄らない場合の経路検索を行い、得られた経路の途中に通る駅全ての駅周辺を検索する。先に経路を決めてから立ち寄る場所を一か所決める手法である。先に立ち寄る場所を決めてから経路を決める手法が理想ではあるが、処理が複雑になる事と検索処理数が多くなる事から、一経路を元に途中の場所を調べる手法を用いた。また、複数個所に立ち寄ってそこで楽しむまたは経験する経路を求める事は稀であり、このような仕様でも問題は無いと考えている。

立ち寄る事を考慮しない起点と終点の経路は、既存の経路検索を用いる。また、途中に通る駅については、駅情報データベースを用いる。途中に立ち寄る場所の候補と選択した「立ち寄る目的」に合うかどうかは、既存のデータベースを用いる。本研究では、飲食店に立ち寄る事を考える。Wikipedia[17] と i タウンページ [18] を用いて、各飲食店の「店名」「外観の写真」「住所」「ジャンル」「食べられる食品（メニュー）と値段」「営業時間」を取得する。

Wikipedia には各鉄道駅毎にページがあり、駅周辺の情報も得られる。例えば鷺別駅 [19] の場合、「駅周辺」という項目から周辺の施設名が得られる。しかし名称のみしか得られないので、各施設に関する詳細情報を得るには Web 検索を用いる必要がある。また、名称からある程度、例えば学校や宗教施設等、語尾や特定の語句等から推測ができる場合がある。名称からの推測によって、検索対象から除外あるいは優先的に行う事ができる。

i タウンページは施設が登録されているデータベースである。駅名を入力すると、駅周辺の施設名が得られる。i タウンページから主に得られる情報は「住所」「カテゴリ」「業種」である。カテゴリは「飲食店」や「ショッピング」等、施設を分類するタグであり、それを元に「立ち寄る目的」に合っているかどうかを決める。「業種」はその施設が取り扱うものであり、例えば飲食店だと「ラーメン店」や「喫茶店」等である。また、一部の店は「食べられる食品（メニュー）と値段」「営業時間」を得ることが可能である。

公式サイトがある場合は、公式サイトを用いた方が信頼度の高い情報が得られる。しかし、名称を知らない状態で公式サイトを見つける事は困難である。そのため、最初に駅周辺の施設が載っているデータベースを用いて、名称を得た後に公式サイトから情報を得る手法を考えた。Wikipedia と i タウンページからでは得るのが難しい「食べられる食品（メニュー）と値段」「営業時間」が得られる。しかしサイト毎に構造が異なるため、各サイト毎に異なる構文解析を用いる必要がある。また、住所やジャンル等一切変わらない情報があり、これらに対しては既存のデータベースの利用でも問題無いと考えられる。

情報を得た後は、出発地から途中に立ち寄る場所、途中に立ち寄る場所から目的地までの経路を検索する。出発地から途中に立ち寄る場所は、入力した時間に出発時刻から経路検索を行う。途中に立ち寄る場所から目的地までの経路は、前の検索結果の到着時刻に、立ち寄るのに最適な時間を加算した時刻を出発時刻として経路検索を行う。現状は立ち寄るのに最適な時間を、立ち寄る目的ごとに決めておく。例えば飲食の場合は 1 時間で、駅前散策の場合は 30 分

である。

経路検索の結果が出力された後は、立ち寄る場所を含めた経路をユーザに表示して選択させる。検索結果を並べる順序の要素として以下が挙げられる。

- 店名を Google で検索した時のヒット数
- 滞在時間の最適な時間との差
- 出発地または目的地からの距離
- 最寄り駅からの距離
- 評判を扱うサイトの評価

他にも考えられるが、本研究では滞在時間の最適な時間との差で考える。検索結果が非常に多くなる事から、全ての立ち寄る場所の候補を表示する事は行わず、上位 10 件をユーザに表示して選ばせる。場所の表示方法については、「施設名」「最寄り駅」「到着と発車時刻」をリストで表示する案と、地図にプロットして表示する案がある。いずれの場合も最小限の情報を表示して、ユーザが選択したら詳細情報及び立ち寄る時刻を表示する考えである。

3.6 途中に立ち寄る場所と混雑度の関係

本研究では、途中に立ち寄る場所の詳細情報を得る事について、混雑度に注目した。その理由は、混雑が原因で予定よりも時間が掛かり、今後の予定が狂う事、また混雑を理由に途中で立ち寄った場所での体験を断念するという事があるからである。そのため、途中に立ち寄る場所を選択する要素になると考えている。しかし、混雑度のみによる判断で途中に立ち寄る場所を選択する事は良くない。ユーザが望む場所や得られる物によっては、多少混んでいても問題無い場合があるからである。

本研究では混雑度を取得するためにツイッターを用いた。その理由として、混雑度は時間によって変化するためである。リアルタイムで情報を得る必要があり、その手段としてツイッターが最適だと考えた。次の第4章では混雑度を含む飲食店の情報を、ツイッターの位置情報を用いて取得を行えるかどうかを検討した。第5章では、実際の店内人数を調査した結果を載せる。第6章では店名のキーワード検索を行い、投稿時にその場に居る事が読み取れるツイートを用いて、店内人数を予測する手法とその結果を載せる。

第4章

SNS を用いる有用性

本章では、途中に立ち寄る場所の候補を得る手段として SNS が利用できないかを述べる。主に混雑度を取得する事を目的としているが、他にどのような情報が得られるかも検討する。

4.1 ツイートから情報を抽出

ツイートからリアルタイムな情報が得られると考えたため、ツイッターを用いて飲食店の情報を得ようと考えた。特に、投稿者がその場に居る事を読み取れるツイートに注目した。例えば、投稿した場所が入っているツイートや、そこで撮った写真が載っているツイートである。このようなツイートからその場の様子や状況、または飲食店や娯楽施設の感想等が書かれている事があり、それらから現地の様子を読み取る事ができる。またその人が飲食店に居る事自体も、有効な情報として利用できないかと考えた。これらを用いて、立ち寄る途中地点の候補に関する情報を取得する事ができないかを考えた。ツイッターはツイートをフォロワーに向けて投稿する以上、よほどの事が無い限り、場所を示さずに状況や感想をツイートする事は殆ど無く、飲食店に居る人が発信した情報を、その飲食店のリアルタイムな情報として活用できると考えた。

また投稿する際に、位置情報を付加する事が可能であり、そこからその場に居る事が読み取れる。位置情報とは、緯度経度 (geocode) と住所 (place) である。緯度経度が付加されているならば、住所情報が付加されている。ここで言う住所とは自治体名である。また、札幌市のように区がある場合はそれも表示される。位置情報はリストから選択して、緯度と経度は端末の GPS を用いて取得した値である。

ツイッターの位置情報を用いてイベントを抽出する研究は行われている。例えば田原琢士らの「Twitter ユーザの活動するローカル地域の発見 [20]」では、ツイートの位置情報を用いて地域の情報を取得する研究を行っている。前田高志ニコラスらの「Twitter 位置情報・テキスト情報を用いた人の移動モデル構築と観光地推薦手法の提案 [21]」では、ツイッターの位置情報を用いて投稿者が向かった先を元にクラス分けを行い、観光に適した場所の推薦を行っている。本章では飲食店における情報の抽出を主に行う。その手法として、札幌駅周辺で発信されたツイートを検索して、それが飲食店かどうかを調べた。また、そのツイートの後に行く場所

や、その前に行った場所にも注目した．場所の特定については，ツイートの内容または写真によってその場にいると判断する．そのツイートの前後のツイートからその人が訪れた場所を特定し，途中に立ち寄る飲食店の候補となる情報を得る．

4.2 位置情報による検索

ツイートにある位置情報は二種類ある．緯度経度で表された情報と自治体や地名等の場所情報である．ツイートをを行う際に，前者は GPS によって付加され，後者は選択肢から選ぶ．本研究においては緯度経度による検索を利用し，場所情報による検索は行わない．理由はリストから投稿場所を選択できるため偽装が可能である事，限られた範囲内からのツイートを検索するにあたって，範囲が広すぎるからである．また，リスト内の場所が指す境界付近におけるツイートの取得が難しく，近くに札幌市北区と東区と中央区の境界がある札幌駅周辺から投稿したツイートを得る場合，場所情報による検索は不適である．よって，緯度経度による検索を行った．

ツイートの検索に緯度と経度と距離を指定して，指定した緯度と経度の距離以内のツイートを検索する事が可能である．札幌駅の緯度と経度 (43.06861111,141.35077777) と周辺 0.5km から投稿されたツイートを取得した．範囲は以下の図 4.1 の通りである．駅から徒歩で周辺の飲食店へ移動する事と，途中に立ち寄る場合はその場で時間はあまり取る余裕が無い予想から，半径 0.5km 以内で十分であると考えた．



図 4.1 1月12日に札幌駅から半径0.5km以内で投稿されたツイート

またチェックインアプリ「Swarm[16]」を用いて札幌駅でチェックインしたツイートも取得して，それらのツイートの後に各投稿者が向かった先を調べた．Swarmとはチェックインアプリで，簡単に居場所を共有することができる．また，待ち合わせや集合等でも活用できるため，広く使われている．その場所に居る事に関する情報の他に，コメントや写真を載せる事ができるため，その場所の状況を知る手段として最適ではないかと考えた．ツイッターと連携して，チェックインしたことをツイッター上で共有することもできる事から，本章ではツイッターに投稿されたチェックイン情報を用いた．

検索に用いたキーワードは「swarmapp.com 札幌駅」と「swarmapp.com さっぽろ駅」である。Swarm によるチェックインは場所名が「i'm at 場所名」と表される。コメントを入れてチェックインした場合は「@場所名」となる。swarmapp.com のリンクが必ずある事を利用して「swarmapp.com 札幌駅」と「swarmapp.com さっぽろ駅」でキーワード検索を行った。札幌駅のチェックインできる場所として他に「札幌駅バスターミナル」や「JR 札幌駅 10 番ホーム」等もあり、その場合も札幌駅に居る事を表すツイートと見なす。直接店名で検索して、飲食店からチェックインしたツイートを直接取得する事は行わない。その理由として、飲食店の数が多く把握しきれない、また飲食店の数だけ検索を行う必要がある事に加えて、飲食店が変わる事があるからである。

図 4.2 は検索結果に該当するツイートと、その後に投稿されたツイートである。一番下のツイートが札幌駅でチェックインしたツイートであり、その後に時計台を経由して麺屋サスケというラーメン屋に行った事が読み取れるツイートがある。また、一番上にそのラーメン屋での感想が書かれている。このようにして、飲食店に居る事を示す前に、駅等人が多く集まる場所に居る事をツイートする事が多いと考え、駅の次に行った先が飲食店であるかどうかを調べ、情報取得として利用できるかどうかを検討する。飲食店以外でも立ち寄るのに適した場所はあるが、ここでは考慮しない事にした。また、該当するツイートまたはその後のツイートに投稿された写真がその場所で撮った写真である事が多いと考えた。その他に、店名を用いずにその場に居る事が分かるツイートを取得する事ができる可能性がある。



図 4.2 駅にいるツイートから飲食店へ向かっている例

4.3 今居るツイートの調査結果

本来は次章の混雑度予測のためのツイート数の追加分として行ったため、検索日は2015年6月15日と7月29日を対象とした。調査した日から10日以上経過しており、Twitter4J[22]を用いて取得が行えず、ツイッター公式の検索を用いた。そのため、各ツイートが発信された緯度と経度を知ることはできない。記録する内容は主に以下の通りである。

- そのツイートが発信された場所
- ツイートされた時刻
- 場所を特定できた要因
- それが飲食店かどうか
- 該当するツイートまたはその後に写真があるか
- そのツイートの後に行く場所と時刻
- そのツイートの前に行った場所と時刻

ツイートが発信された場所は主に文章の内容または写真を元に特定する。札幌駅から投稿されたツイートは「札幌駅から」、札幌駅周辺から投稿されたツイートは「飲食店から」、それ以外の場所から投稿されたツイートは「上記以外」に分類する。特定できない場合は「判別不能」とし、前後のツイートを確認できないツイートや、明らかに宣伝目的で投稿されているツイートは「無効」とする。

場所を特定できた要因は、Swarm と Instagram[23] から特定できたツイートが多い事から、それらから特定できた場合とその他に分ける。そのツイートの後に確認できる行った場所は、その日の内かつ4時間以内であれば直後でなくても良い。例えば、札幌駅に行ったツイートの後に隣の駅へ行ったツイートがあり、その後に温泉に行った記録がある場合は、温泉へ行った事でも良いとする。但し、そのツイートの前に行った場所は、その日の内かつ直後に確認できる場所を記録する。写真のみで投稿場所が特定できない場合は特定不能として数える。

表4.1は得られたツイートの前後に関する結果をまとめたものである。緯度経度による検索で、場所を特定できた要因の多くはSwarmからによるツイートである。それに次いで、Instagramによるツイートや文章や写真から投稿場所を特定できたツイートがある。SwarmやInstagramによるツイートの緯度経度が付加されることがあるため、このような結果になった。そのため、緯度経度の検索結果とSwarmからのツイートを検索した結果に同じツイートがある。しかし、表4.1(b)のSwarmの札幌駅のチェックイン数と、(a)のSwarmからのツイート数が一致しない事から、全てのSwarmからのツイートの緯度経度があるとは限らない。また、Swarmからのツイートの検索結果は「札幌駅」の他に、飲食店の「札幌駅前店」と「新札幌駅」からのチェックインが見られた。

ツイートされた場所を追いかけると、旅行中である事が分かる例もある。例えば札幌駅に行った後に新千歳空港へ行き本州へ戻る、または旅行や講演等で札幌へ行く、札幌の家に戻ってくるツイートがあった。これらの事から、旅行者が推薦する場所を示すことに使えると考え

られる。また、札幌駅にチェックインした後に周辺の飲食店へ行くと予想していたが、実際は周辺の飲食店へ行った後に札幌駅にチェックインするというのがあり、そこから遠方へ行く内容もあった。

表 4.1 札幌駅周辺で発信されたツイートの調査結果

	2015年6月15日		2015年7月29日	
	(a) 緯度経度から	(b) Swarm から	緯度経度から	Swarm から
全体のツイート	45件	41件	55件	39件
札幌駅から	10件	29件	13件	20件
飲食店から	13件	5件	14件	3件
上記以外	8件	5件	11件	3件
特定不能	5件	0件	10件	0件
無効	9件	2件	7件	0件
場所を 特定できた 要因	Swarm	26件	27件	
	Instagram	4件	7件	
	その他	1件	5件	
それまたは後に写真がある	9件	6件	19件	7件
追跡可能	17件	26件	16件	30件
1km 以内	8件	11件	12件	13件
2時間以内	11件	19件	13件	24件
飲食店へ行く	5件	4件	6件	4件
特定不明	2件	0件	3件	0件
直前の場所が特定できる	16件	26件	24件	32件
1km 以内	6件	12件	14件	14件
2時間以内	14件	21件	19件	25件
飲食店へ行った	0件	5件	4件	5件

今回の調査では飲食店でツイートしても、以下のような場合は反映されない。

- 位置情報がオフになっている
- 店名等特定できる情報が無い
- 駅で Swarm のチェックインをしない、またはツイッターで共有しない

位置情報をオフにする事については、位置情報をオフにする事を推奨する記事が多い事から、自宅が特定される可能性がある事が話題になったからであろう。それにより、緯度経度を付加したツイートをあまりしない人が多いと考えられる。また同様に、むやみに今居る事を知らせるを控える傾向があるのではないかと考えられる。表 4.1 では札幌駅周辺の飲食店からのツイートが札幌駅でのツイートよりも多い。しかし、ここでいう飲食店は札幌駅の周辺 0.5km 以内の飲食店全てと広範囲である。駅から発信するツイートは多く読み取れるであろう。その

理由として、駅の場合は人の通りが多く簡単にはアカウントの特定がされないと考えている人が多いからであろう。そのためか、駅で投稿したツイートの前後に向かった先の情報が、駅以外の場所が一切無い場合が多い。同様に店名をツイートに出さない事が多いと考えられる。

調査したツイートの特徴をまとめると、感想や状況が文章で書かれているツイートは殆どなく、大抵はチェックインして終わりである。場所を特定できない要因として、文章量の不足が挙げられる。緯度と経度が分かれば良いが、10日以上前のデータを用いたため、取得する事が出来なかった。一方で、写真で状況を発信している事はあった。特に、Swarm のツイートの後に写真がある場合は、その場所の風景や商品の写真である事が多く、ツイートした場所またはこれから向かう場所で撮った写真を投稿する事が多いと考えられる。この調査においては、店名を出さずに食品の画像のみが載っているツイートは無かった。しかし、写真による場所の特定ができない事もある。投稿された町中の画像から、GoogleEarth を用いて撮影場所を特定することは可能であるが、写真の中のわずかな情報から推測する必要があるため、自動化は不可能だと考えられる。商品の写真のみが載っている場合は、メニューや案内に載っている写真と照らし合わせる作業が必要だが、今回の調査でそのような事は無く、その店に関する情報として利用できる可能性が高いと考えられる。

写真以外に有益な情報は殆ど無いが、ツイートは多くて140文字であるが、文字の入力には時間が掛かるため、時間が無いまたは混雑等で食事に時間を掛けられない場合はツイートを控える事が有り得る。また、商品の写真がある場合がその場で写真を撮る必要がある。この事から居場所を知らせるツイートは、その店がツイートできる余裕があるほど混んではいけない情報になると考えられる。該当するツイートまたはその前後に飲食店へ行く事が読み取れるツイート数は

- 緯度経度の6月15日...16件(全体の35.6%)
- 緯度経度の7月29日...20件(全体の36.4%)
- Swarm の6月15日...12件(全体の29.3%)
- Swarm の7月29日...12件(全体の20.8%)

であり、半数以下ではあるがどのような飲食店があるかを知るには適していると考えられる。

第 5 章

混雑度の調査

本章では、ある飲食店を対象に人手で実際に測った混雑度と今居るツイート数の調査方法、その結果について述べる。混雑度の調査対象として、札幌駅エスタ 10 階の「ら～めん共和国」を選んだ。「ら～めん共和国」はフードテーマパークで 8 つの店で構成されており、この調査では 8 つの店全て合計した値を用いる。総座席数は 250 席である。

混雑度を測るうえで、以下を定義する。

- 座席数...店内にある座席数
- 店内人数...店内の座席に座っている人数
- 座席占有数...空いていない座席

座席占有数は座席配置の都合上、座っている人が居ないのに座れない席である。例えば、4 人テーブルで一人で座っている場合は人数は 1、座席占有数は 4 となる。その他に、手荷物が置かれている席も座席占有数として数える。また、店内に座っている人のみを数え、列に並んでいる人は数えない。8 つの店全てを調査した値を用いており調査時刻は一瞬ではないため、調査開始と終了時刻の中間値を調査時刻とする。

調査日は 6 月 10 日、6 月 12 日、6 月 14 日、6 月 15 日、7 月 29 日である。都合上、各調査日とも営業時間全てのデータを得る事ができなかった。営業時間は 11 時から 22 時であるため、11 時と 22 時の店内人数と占有数を 0 とする。実際に調査した店内人数と座席占有数を以下の図 5.1 から図 5.5 に載せる。

各グラフから店内人数及び占有数については、各日とも 12 時から 13 時と 19 時から 20 時 30 分の間に店内人数が最も多く、200 席近くとなっている。これは、昼食と夕食時の影響であると考えられる。店内人数と占有数両方が総座席数の 250 に達している事は無いが、実際の調査において 8 つのうち幾つかの店が混んでいる時があった。また各店に空いている席があるにも関わらず満員となり、行列が発生してる店があった。この事から、店内人数及び占有数が最大数にならなくても行列が発生して混雑すると考えられる。

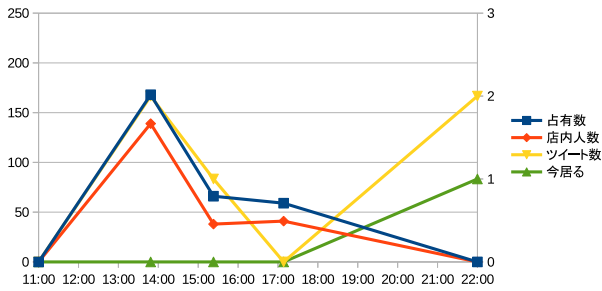


図 5.1 6月10日の店内人数とツイート数

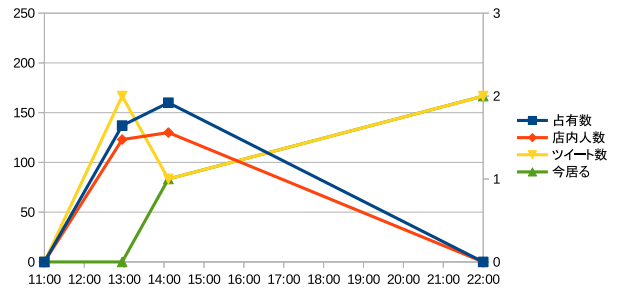


図 5.2 6月12日の店内人数とツイート数

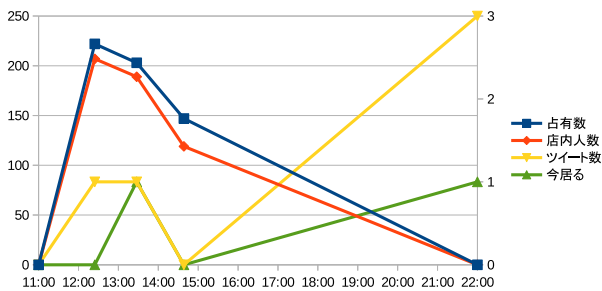


図 5.3 6月14日の店内人数とツイート数

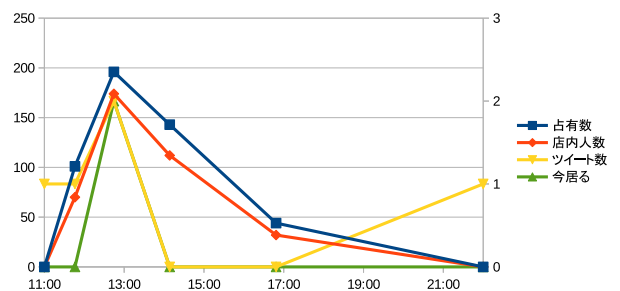


図 5.4 6月15日の店内人数とツイート数

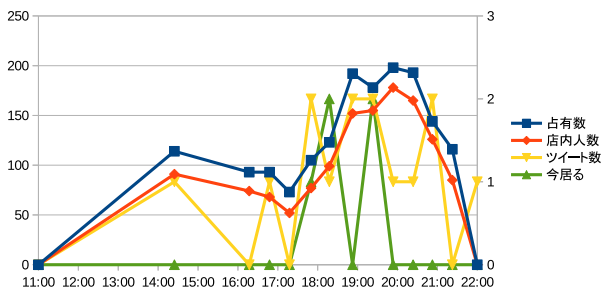


図 5.5 7月29日の店内人数とツイート数

第6章

混雑度の予測

本章では混雑度を予測する手法を説明する。

本研究ではツイッターを用いて混雑度を予測する事を考えた。混雑は時間毎に変化するため、リアルタイムで現在の情報が得られる事が望ましい。その理由として、毎日混んでいる事があっても混雑度は一定でないため、空いている時間帯があり、このわずかな時間内の変化を捉える必要があるからである。また、ある程度パターンや法則性はあるものの、混雑度は時刻、天候、イベントの有無等、様々な要因が関わる上これらもリアルタイムで変化する事から、過去のデータと同じになる事は無いと考えられる。

現在実用化されている混雑度測定はハードウェアを用いる手法が主であり、混雑度を予測するに至って、設備を導入する必要がある。そのため、設備の導入する手間が掛かる事と、導入されていない場所における混雑度測定はできないという問題がある。そこで、ツイッターのユーザが現在地を知らせる投稿を利用しようと考えた。ツイッターはインターネット環境があれば投稿が可能であるため、特別な設備の導入は不要である。また、ユーザがこの場に居る事やその場の状況を投稿する事があり、これらからリアルタイムの情報が得られると考えた。

6.1 ツイート数の取得方法

ツイッターを用いて予測するにあたって、今この場に居る事が読み取れるツイートに注目した。ツイートの内容や写真から推測する手法もあるが、第4.3節の結果から文章による飲食店に関するの情報抽出が難しい。また写真は取得できるものの多くは商品の写真であり、店内の様子を表す写真は無いため、写真による混雑度が難しい。以上の事から、ツイートの内容ではなくツイートの総数を用いた。店内人数が多いとその場所から発信されるツイート数が多くなり、今居る事を示すツイートの数もそれに伴って多くなるという考えである。本当に今そこに居るという保証は無いが、確かめる方法は無いため、実際にその場に居ると仮定する。

今居る事が分かるツイートとは、例えば「～なう」や図6.1のように「I'm at」「@(場所名)」と書かれたツイートである。それ以外にも図6.2のように読み取れるツイートがある。この他に図6.3のように写真の内容や付加した位置情報から場所の推定が可能なツイートも該当する。また、「行ってくる」等これから行く事が読み取れるツイートもあるが、多くは来る

時刻が書かれていない事が多く、その場合はツイート時刻から 30 分後にその場に居ると仮定する。その他、図 6.4 のようにツイートの返信や、その後のツイートからその場に居るまたはこれから行く内容だと判断できる場合がある。また、図 6.5 のように店名を用いていないが写真から読み取れるツイートや、図 6.6 のように写真のみのツイートでもその場に居る事が読み取れることがある。他にこれらのパターンに当てはまらないでその場に居る事が読み取れるツイートが多く存在しているため、判断基準は明確になっていない。そのため、今居る事が読み取れるかどうかは、著者の判断によって行われた。

ツイートの内容を読み取って例えば店の商品の評判、混んでいるまたは空いている、おすすめの商品、限定のイベントや商品、公式からの案内等の情報を得る事ができる。しかし第 4 章の調査から文章で様子が読み取れるツイートが殆ど無く、ツイートの内容から情報を抽出する事は非効率だと考えた。また、過去の話である事も多く、今の情報を得られない事もある。その他に、ツイートの内容からどのような情報が書かれているかを判別し、ツイート全体のどの部分が得る情報なのかを抽出する必要がある。

調査において、Twitter4J[22] を用いて Java 上で TwitterAPI を動かしてツイートの内容、投稿時刻を取得した。検索に用いたキーワードは、「ら～めん共和国」の他によく見られる誤表記である「ラ～メン共和国」「らーめん共和国」「ラーメン共和国」を用いた。ラーメン共和国内の各店名を用いた検索は行っていない。絶対数が少ない事もあるが、支店が判別不可能のツイートがあり、調査には向かなかった。緯度経度情報を用いた検索については、ら～めん共和国が屋内である事に加えて、飲食店が密集している場所にあるため、周辺の施設からのツイートと判断が付かない事があり行っていない。リツイートは結果から除外する。収集したツイートのうち、実際の店内人数を測った時刻から 30 分前の結果と、毎時 0 分と 30 分それぞれの 30 分前のツイート総数をツイート数とする。



図 6.1 Swarm から投稿されたツイートの例



図 6.2 文章で今居る事が読み取れるツイートの例



図 6.3 写真が載っている今居る事が読み取れるツイートの例



図 6.4 会話の内容から今居る事が読み取れるツイートの例



図 6.5 店名を用いなくて今居る事が読み取れるツイートの例



図 6.6 写真のみの今居る事が読み取れるツイートの例

6.2 ツイート数の結果と考察

ツイート数の結果は図 5.1 から図 5.5 に、実際の店内人数と同時に載せる。同じ日の店内人数とツイート数を比較すると、それぞれの値がピークになっている時刻が一致していない事がある。例えば 6 月 10 日の結果（図 5.1）からは、店内人数のピークは 14 時であるが、ツイート数のピークは 22 時である。6 月 14 日（図 5.3）と 7 月 29 日（図 5.5）も同様に、店内人数のピーク時に今居るツイート数が無い。このように、最も店内人数が多い時に今居るツイート数が多いとは限らない。その一方で 6 月 12 日（図 5.2）と 6 月 15 日（図 5.4）のように、店内人数と今居るツイート数のピークが 13 時と一致しているがある場合があり、相関が無いとは言えない。また、得られた 30 分前までのツイート数は最大 3 件であり、今居るツイートも最大 2 件と少ない。店内人数と占有数の変化率と各ツイート数の変化率に大きな差があるため、正確な予測は困難であると考えられる。

調査時刻から 30 分前のツイート数が最大 3 件と少ない理由としては、店名が載っているツイートのみを対象としたからであると考えられる。ツイートの検索手法にキーワード検索があり、この検索手法から今居る事を示すツイートを得るには店名を用いるのが効率が良い。しかし、実際にその場に居る事を示すツイートには写真のみを載せる等文字を用いない、または「着いた」等店名を用いず一単語や動詞のみのツイートが多く、このようなツイートを既存のキーワード検索のみで得るのは難しい。また、ツイートを遡って調査する事が大変であるため、タイムラインやリプライによる判断は考慮しなかった。

6.3 ツイート数をそのまま用いた単回帰分析

各調査日の全調査時間における今居るツイート数と占有数及び店内人数の相関係数を表 6.1 の (a) に載せる．表の結果からは，著者の時間の都合により調査した時刻に空きがある事や全時間について調査できなかった部分もあり，相関は殆ど無い．理由として，店内人数が最も多い時と今居るツイート数が最も多い時が同じとは限らない事が多い．これは混んでいる時にツイートをする余裕が無いからであると考えられる．その一方で，図 5.4 を除くと今居るツイートよりも店名が書かれているツイート全体を用いた方が良い結果になる場合がある．

また，図 5.1 から図 5.5 の 5 つ全ての 22 時は営業終了時刻で店内人数と占有数が 0 にも関わらずツイート数がある．営業終了時刻 30 分前でも人は居るため，ツイート数が 0 以上になる事はあり得るが，営業終了時間以降の店内人数と占有数の両方が 0 だと分かる事，この時間に店に行く必要は無い事から，終了時間のツイート数を 0 と見なしても良いだろう．同じ事は営業開始時刻の 11 時以前にも言える．以上から，22 時のツイート数を 0 とした時の相関係数の値も表 6.1 の (b) に載せる．

表 6.1 より，閉店時のツイートを 0 にする事によって計算不能となった 6 月 10 日を除くと，(a) より相関係数が大きくなった．6 月 10 日の (b) は 22 時のデータを除外した事により，全ての時刻における今居るツイート数が 0 であるため，計算不可となっている．そのため，全ての時間を用いる (a) の結果よりも，(b) の結果を用いた方が良いと考えられる．但し，6 月 10 日と 6 月 12 日，6 月 14 日は両端を含めた調査点が 4~5 と少ないため，相関係数が異常に上がったと考えられる．

表 6.1 ツイート数と店内人数及び占有数の相関係数

		(a) 全時間を対象とした場合		(b) 閉店時のツイートを 0 件と見なした場合	
		占有数	店内人数	占有数	店内人数
6 月 10 日 (図 5.1)	今居る	-0.4767	-0.4285	計算不可	計算不可
	全体	0.3965	0.4307	0.9200	0.9101
6 月 12 日 (図 5.2)	今居る	-0.2533	-0.2846	0.6628	-0.6088
	全体	0.2533	0.2846	0.8527	0.8872
6 月 14 日 (図 5.3)	今居る	-0.1090	-0.0779	0.4586	0.4827
	全体	-0.2778	-0.2439	0.8291	0.8707
6 月 15 日 (図 5.4)	今居る	0.4485	0.5352	0.7067	0.7787
	全体	0.2980	0.3657	0.5219	0.5674
7 月 29 日 (図 5.5)	今居る	0.2064	0.2072	0.2064	0.2072
	全体	0.6065	0.5955	0.6580	0.6493

店名を単に含むだけのツイートを用いるか店名を含む今居る事が分かるツイートのみを用いるかについては、6月15日は今居るツイートを用いた方が、7月29日は全体を用いた方が相関係数が大きくなっており、表 6.1 のみではどちらが良いとは言えない。但し、7月29日は調査点が多い事と間隔が狭い事を考えると、ツイート全体を用いた方が良い結果が得られると考えられる。

図 6.7 から図 6.11 は単回帰分析の結果である。表 6.1 の (a) で、今居るツイート数と店内人数の相関係数が最も高かった 6月15日の結果を用いて、単回帰分析によって2つのデータ間の近似式を求めた。得られた近似式は以下の通りである。

$$\{ \text{店内人数} \} = 65.6 \times \{ \text{今居るツイート数} \} + 42.8$$

この近似式に基づいて、各調査日の30分毎の今居るツイート数を当てはめて求めた店内人数の予測値と、調査した店内人数とを比較した図である。図 6.9 と図 6.11 からは、実際の店内人数の最大値と単回帰分析で求めた近似式による最大値の時刻が一致していない。最も混んでいる時にツイート数が最も多いとは限らないからであると考えられる。また、図 6.7 と図 6.8 と図 6.10 は後ろ半分が調査結果に似ているが、実際は夜間の調査が全くできずデータが無いため調査結果の通りにはならず、図 6.11 のように 19 時頃に店内人数が多くなるであろう。

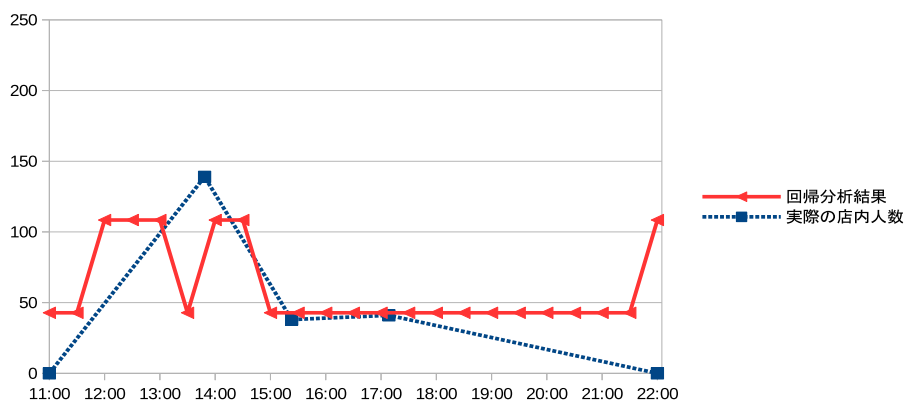


図 6.7 6月10日の単回帰分析に基づく店内人数と実際の店内人数

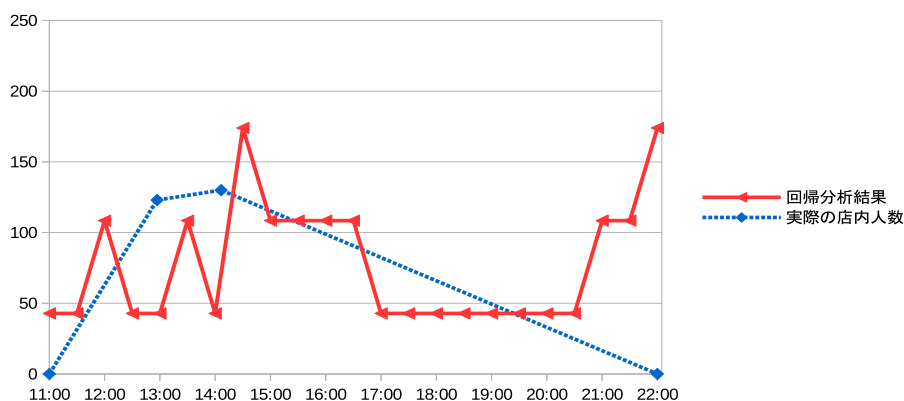


図 6.8 6月12日の単回帰分析に基づく店内人数と実際の店内人数

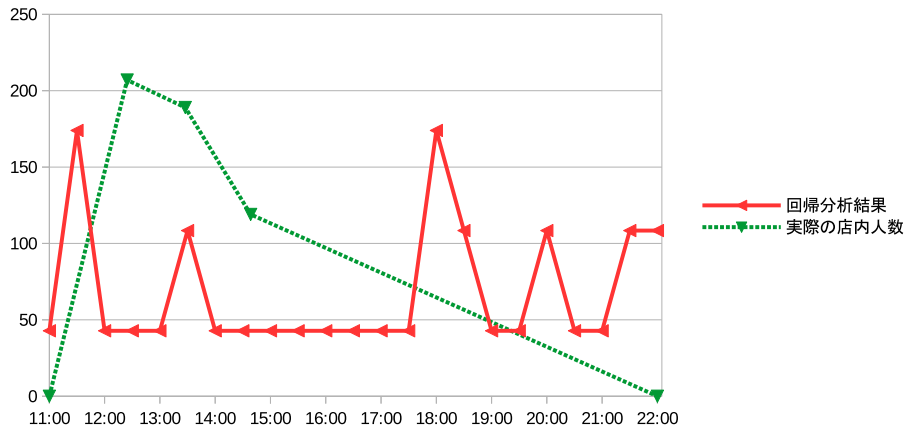


図 6.9 6 月 14 日の単回帰分析に基づく店内人数と実際の店内人数

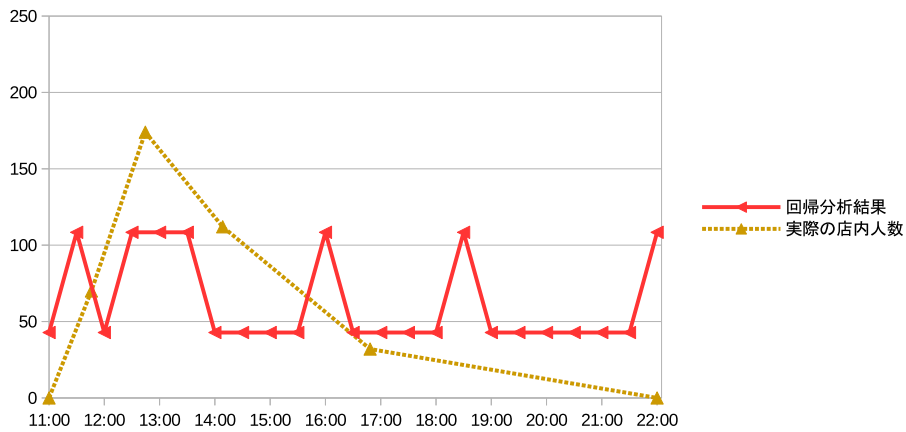


図 6.10 6 月 15 日の単回帰分析に基づく店内人数と実際の店内人数

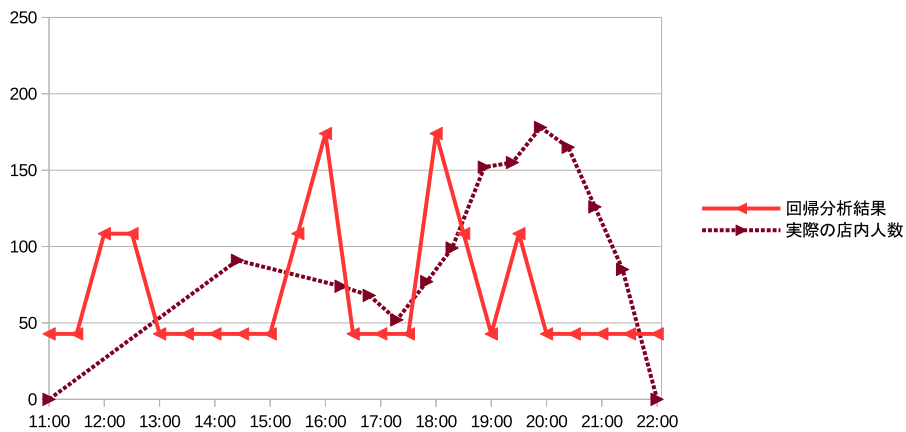


図 6.11 7 月 29 日の単回帰分析に基づく店内人数と実際の店内人数

6.4 残存効果の適用

実際の店内人数が多段階で変化するのに対し、ツイート数が最大2つしかないため、予測店内人数も3段階となっている事が問題である。前の時間のツイート数の残存効果がある程度影響を与えると考え、それを足した値と比較した。このようにして、ツイート数を多段階に変化させて実際の店内人数に単回帰分析を行う。

今居るツイート数に以下の式を適用する。

$$Twn(t) = a \cdot Tw(t) + b \cdot Tw(t - \Delta t) + c \cdot Tw(t - 2\Delta t)$$

$Twn(t)$ は前の時間のツイート数の残存効果を適用したツイート数で、 $Tw(t)$ は適用する前のツイート数である。 a, b, c は重みであり、それぞれ $a+b+c=1, 0 \leq a \leq 1, 0 \leq b \leq 1, 0 \leq c \leq 1$ であり0.1刻みとする。 $Tw(t - \Delta t)$ は時刻 t から Δt 分前における今居るツイート数である。本章ではツイート数の残存効果と呼ぶ。

6月15日の今居る事が分かるツイート数と、7月29日の今居る事が分かるツイート数を対象に残存効果の適用を行った。実際の店内人数との相関係数と重み a, b, c の各値を表6.2に示す。一番上の $a=1, b=0, c=0$ の結果が表6.1の6月15日と7月29日の結果に対応している。二番目から六番目は相関係数を高い順に5つ並べた結果で、一番下はF値 ($F = (2 \times 6月15日の相関係数 \times 7月29日の相関係数) / (6月15日の相関係数 + 7月29日の相関係数)$) が最も大きい時の係数の結果である。F値の最大値は0.6826である。

表6.2 前の時間のツイート数の効果を適用したツイート数の相関係数

6月15日の係数				7月29日の係数			
現時刻 a	30分前 b	60分前 c	相関係数	現時刻 a	30分前 b	60分前 c	相関係数
1	0	0	0.5352	1	0	0	0.2072
0.3	0.2	0.5	0.8547	0.2	0.4	0.4	0.5856
0.3	0.3	0.4	0.8530	0.3	0.4	0.3	0.5773
0.4	0.2	0.4	0.7988	0.2	0.5	0.3	0.5747
0.3	0.1	0.6	0.7917	0.1	0.5	0.4	0.5745
0.4	0.1	0.5	0.7842	0.1	0.4	0.5	0.5728
0.3	0.3	0.4	0.8530	0.3	0.3	0.4	0.5689

両日とも、前の時刻のツイート数の残存効果を適用する場合と比べて、相関係数が上がっている事が分かる。また一つを除いて $a < c$ となっており、今の時刻のツイートよりも1時間前のツイートが重視される結果となった。しかし今回の調査はラーメン屋を対象にしたため、1時間も居続ける事は無いだろう。今居るツイートを発信した人は、発信してから1時間後には既にその場に居ないと考えられる。そのため、1時間前のツイート数と店内人数の関係は不明である。

F 値が最も高い $a = 0.3, b = 0.3, c = 0.4$ を用いて残存効果の適用を行った。 $Tw_n(t) = 0.3 \cdot Tw(t) + 0.3 \cdot Tw(t - 0.5) + 0.4 \cdot Tw(t - 1)$ を用いたツイート数を元に、6月15日の結果を用いて単回帰分析を行った。残存効果を適用した結果(緑)と残存効果を適用しなかった結果(赤)と実際の店内人数(青)を図 6.12 から図 6.16 に載せる。得られた近似式は

$$\{ \text{店内人数} \} = 183.6 \times \{ \text{残存効果を適用した今居るツイート数} \} - 5.7$$

である。

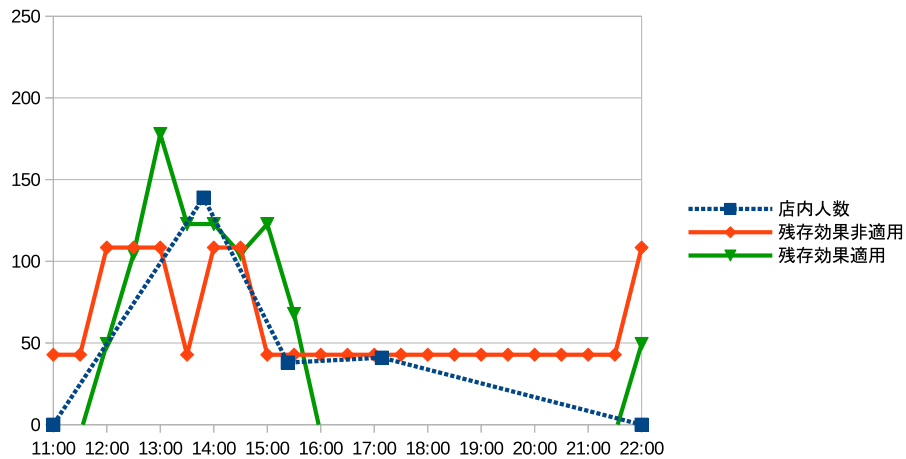


図 6.12 6月10日の残存効果適用による予測店内人数

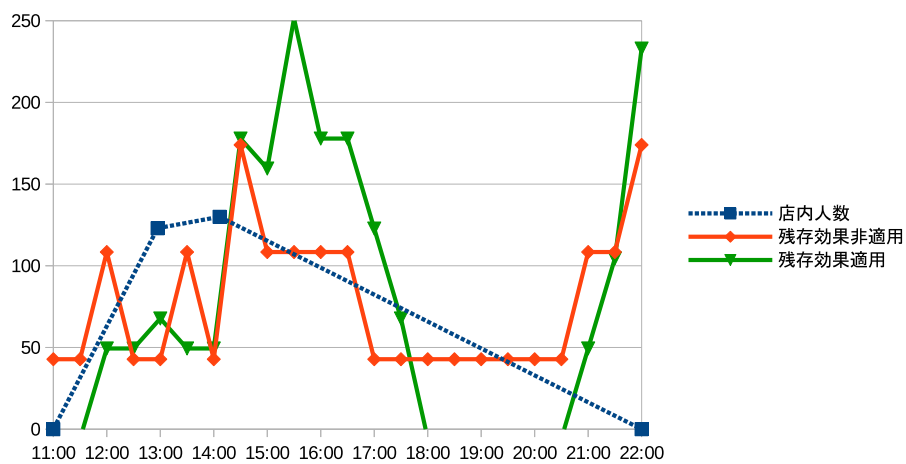


図 6.13 6月12日の残存効果適用による予測店内人数

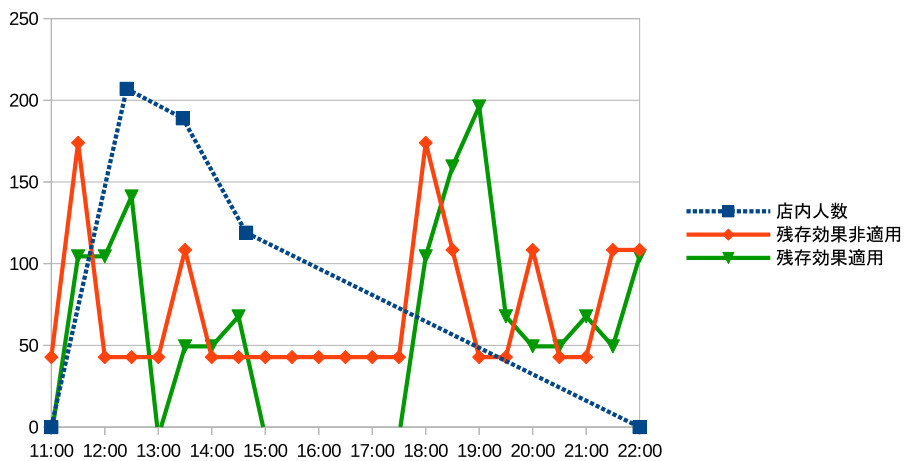


図 6.14 6月14日の残存効果適用による予測店内人数

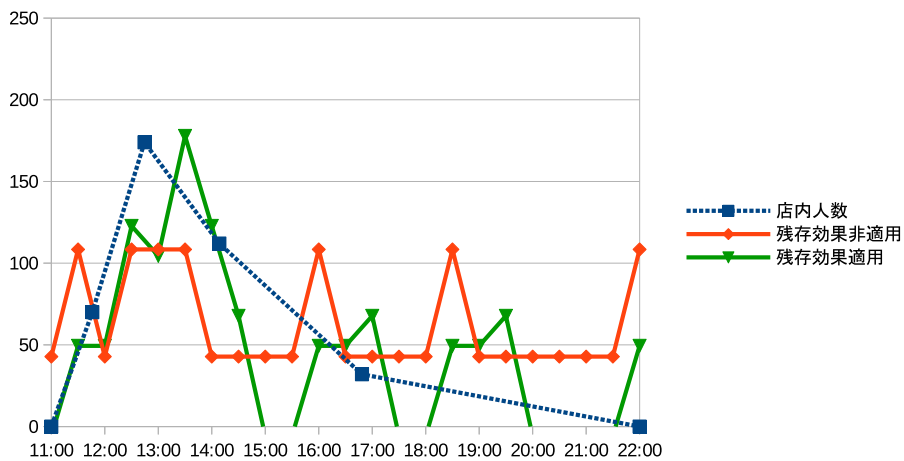


図 6.15 6月15日の残存効果適用による予測店内人数

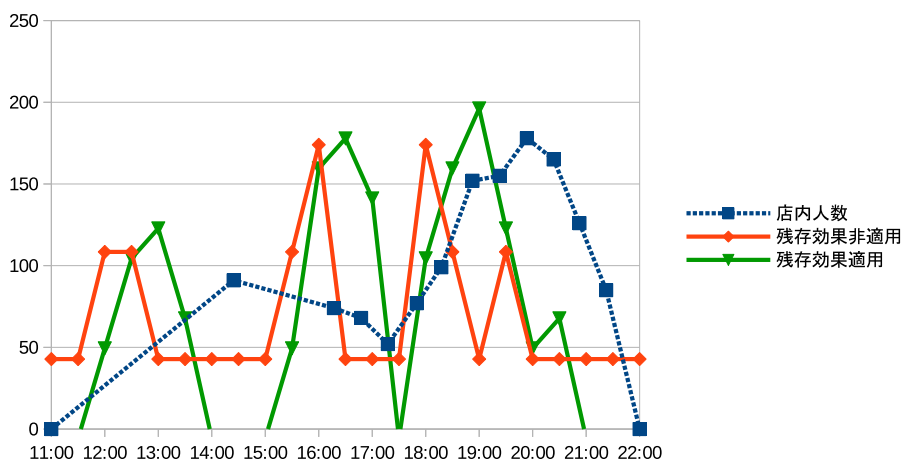


図 6.16 7月29日の残存効果適用による予測店内人数

残存効果を適用する事で、図 6.15 の 11 時から 14 時や図 6.16 の 17 時から 19 時のように、適用前と比べて予測店内人数が細かく変化するようになった。このように、変化が緩やかになった事で予測店内人数が実際の店内人数に近づいたことが読み取れる。しかし、図 6.13 の 14 時から 18 時のように、あまり混まないと予想される時間帯で座席数の 250 を超えている結果や、図 6.16 の 14 時から 17 時のように実際の店内人数が横ばいになっている状態で、予測店内人数が急激に変化するなど、実際の値と異なる部分がある。また、どの日の結果も店内人数が最大の時に、予測店内人数が実際の店内人数を下回っている。その一方で、予測値が実際の店内人数の最大値を超える事があり、図 6.14 を除いて、残存効果適用の予測店内人数が実際の店内人数の最大値を取る時刻の約 30 分前か後に実際の店内人数の最大値を超えている。実際の店内人数が最大値の前に予測店内人数が下がる場合は、それを元にピーク時の予測が可能であると考えられるが、そうではない場合がある。また、予想店内人数がマイナスの値を取る問題がある。

6.5 時間帯を分けて残存効果を適用

図 6.16 から、18 時から 22 時における予測店内人数と実際の店内人数は似たような形になっているが、14 時から 18 時では実際の店内人数が横ばいで、予測店内人数が上に凸となっており、大きく形が異なっている。この事から時間帯毎に適切な値が異なるのではないかと考えた。調査した店内人数には空白時間帯があるもの、12 時頃を中心に極大を取る、15 時から 18 時は横ばいとなり、19 時から 20 時に極大を取ると予想できる。この事から、11 時～15 時、15 時～18 時、18 時～22 時と分けて考えた。

調査点の都合上、11 時～15 時は 6 月 14 日と 6 月 15 日の店内人数と今居るツイート数の相関係数のうち、F 値が高い時の係数、15 時～18 時と 18 時～22 時は 7 月 29 日のデータを用いて、最も相関係数の高い値を用いる。その結果、11 時～15 時は $a = 0.3, b = 0.2, c = 0.5$ 、15 時～18 時は $a = 0.8, b = 0.2, c = 0$ 、18 時～22 時は $a = 0, b = 0.5, c = 0.5$ となった。よって、残存効果を適用するツイート数の数式は以下の通りである。

$$Twn(t) = 0.3 \cdot Tw(t) + 0.2 \cdot Tw(t - 0.5) + 0.5 \cdot Tw(t - 1) \quad (11 \leq t < 15) \quad (6.1)$$

$$Twn(t) = 0.8 \cdot Tw(t) + 0.2 \cdot Tw(t - 0.5) \quad (15 \leq t < 18) \quad (6.2)$$

$$Twn(t) = 0.5 \cdot Tw(t - 0.5) + 0.5 \cdot Tw(t - 1) \quad (18 \leq t < 22) \quad (6.3)$$

11 時から 15 時までは 1 時間後のツイート数が重視される。15 時から 18 時までは、現時刻のツイート数が重視されて、1 時間後のツイートは使わない。18 時から 22 時までは、現時刻のツイート数は一切使われない事となった。

残存効果を適用して時間帯を分けて単回帰分析を行った結果 (緑) と今居るツイート数をそのまま単回帰分析を行った結果 (赤) と実際の店内人数 (青) を図 6.17 から図 6.21 に載せる。得られた近似式は

$$\{ \text{店内人数} \} = 167.8 \times \{ \text{残存効果を適用した今居るツイート数} \} + 17.7 \dots (11 \leq t < 15)$$

$$\{ \text{店内人数} \} = 0.8 \times \{ \text{残存効果を適用した今居るツイート数} \} + 67.7 \dots (15 \leq t < 18)$$

$$\{ \text{店内人数} \} = 73.9 \times \{ \text{残存効果を適用した今居るツイート数} \} + 73.8 \dots (18 \leq t < 22)$$

である。

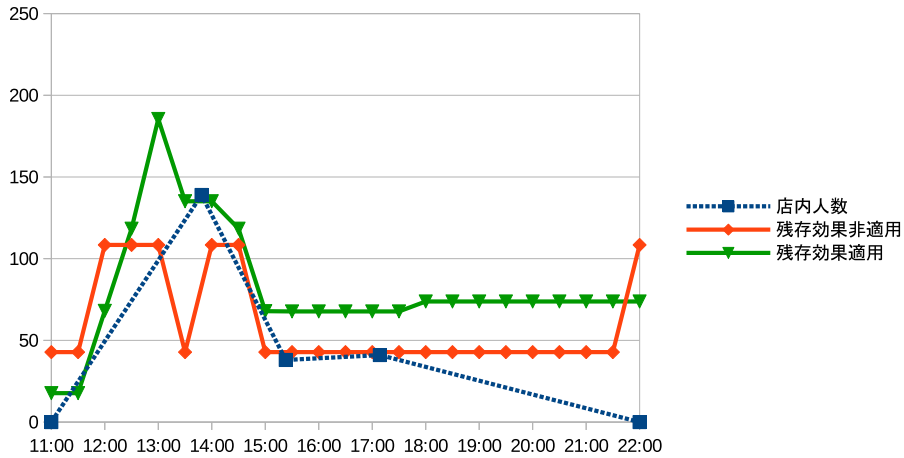


図 6.17 6 月 10 日の時間帯を分けた予測店内人数

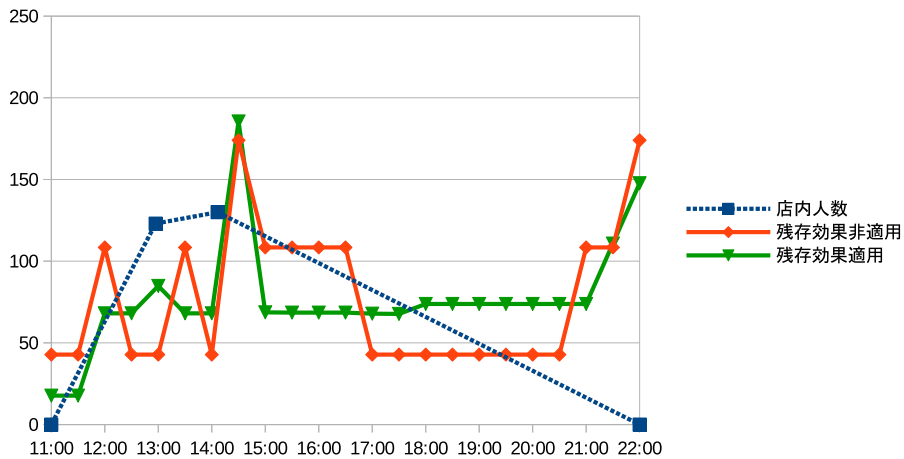


図 6.18 6 月 12 日の時間帯を分けた予測店内人数

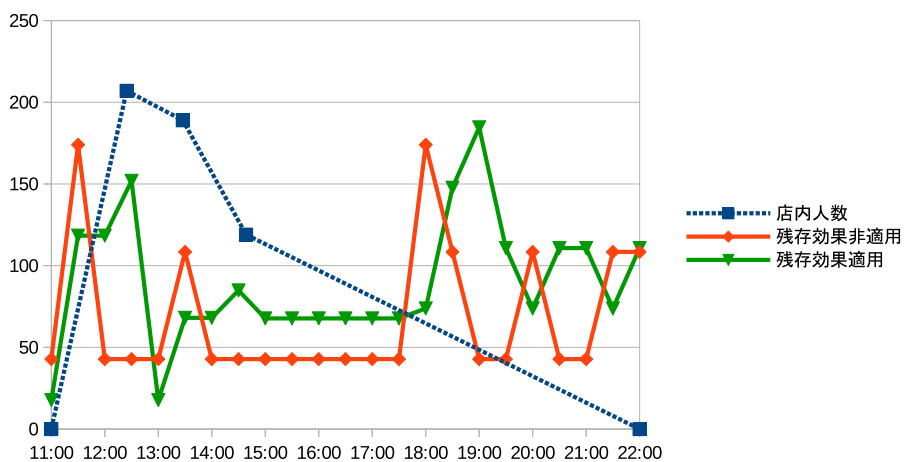


図 6.19 6 月 14 日の時間帯を分けた予測店内人数

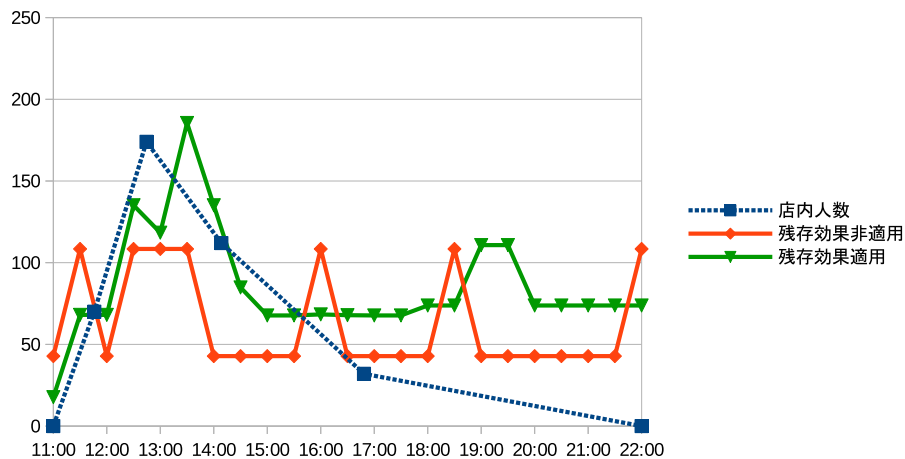


図 6.20 6 月 15 日の時間帯を分けた予測店内人数

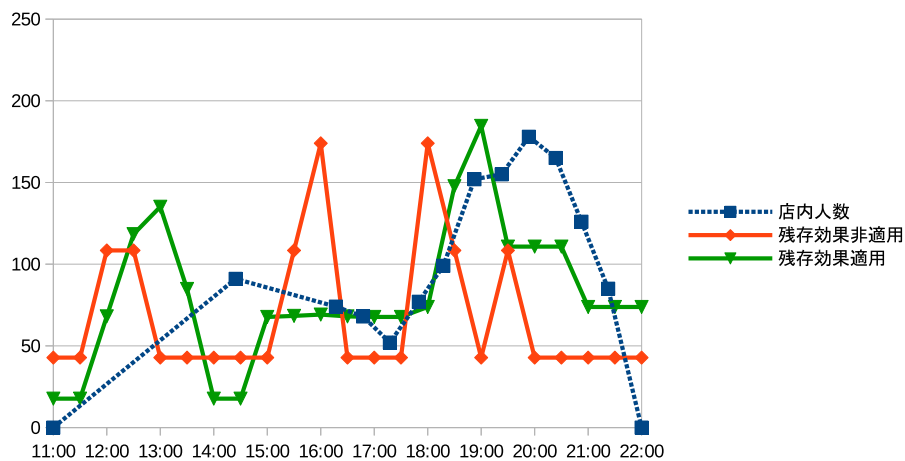


図 6.21 7 月 29 日の時間帯を分けた予測店内人数

15 時から 18 時までは時間帯を分けて考えなかった図 6.17 から図 6.21 と比べると、ほぼ実際の店内人数と似たような結果となった。ツイート数が店内人数に殆ど影響を与えない状態である。残存効果を適用したツイート数が 166 以上ある場合に、店内人数が全体の 75% に当たる 200 人いるという計算であり、この時間帯において混雑を予測する事は難しいと考えられる。普段は混まない時間帯であろうが、突然人が多く来る可能性も無いとは言えないため、劇的に混む事を予測できた方が良さだろう。図 6.17 からは、15 時までは実際の店内人数とほぼ同じ形となっている。しかし 11 時の次の調査時刻が 14 時であり、予測の結果が正確かどうかは不明である。図 6.18 と図 6.19 は店内人数が最大値の時に予測値が下回ったままである。11 時から 12 時における予測の変化率と実際の変化率は近い値を取っているように見えるが、その後は予測値が横ばいあるいは逆に下がっている。

第7章

結論

旅行や出張等で遠くへ移動を行う際において、経路検索エンジンを使う機会は多い。多くの経路検索エンジンが最短経路や最安値、乗り換え数に重点を置いているが、途中に通る場所はあまり考慮していない事が多い。本研究では、途中に立ち寄る事を考慮した経路検索を提案した。通る経路の途中でイベントや食事、絶景等途中経過で起こる事や見られる物から経路を選択する必要があると考えたからである。途中に立ち寄れる場所をユーザに提示して立ち寄る場所を選ぶシステムである。そのために、途中地点を選ぶ基準を複数挙げて、それらをどう取得するかを提案した。本研究では主に飲食店に立ち寄る事を中心に考えて、立ち寄る場所を選ぶ要素として混雑に注目した。理由は混雑によって旅行計画が狂う事、また立ち寄る場所を楽しめなくなる事があるからである。混雑度を取得する手法については、SNSであるツイッターを用いた。

ツイッターを用いて混雑度を取得する前に、SNSを用いて飲食店の情報を得る事ができないかどうかを確認した。ツイッターの緯度経度検索と Swarm[16] のチェックインによる投稿から、札幌駅周辺で投稿されたツイートを検索した。その結果、札幌駅周辺の飲食店から投稿されたツイートの他に、これから向かうまたはその前に行った飲食店から投稿されたツイートを取得できた。しかし、写真による情報取得は可能である一方で、感想や状況を文章で得る事は難しい、この事から、混んでいる事をツイートの内容から得るのは難しいと考え、今その場に居る事を表すツイートの数を用いて混雑度の予測を行った。

今居る事が分かるツイートを取得する手法として店名でキーワード検索を行い、得られたツイートの内容からその場に居る事が分かるツイートを選ぶ事で得る。今回の調査においては、内容を一切考慮せずに食事前にツイートしたという前提で考え、調査時間から30分前のツイート数を用いた。今居る事が読み取れるツイート数と、実際に調査した店内人数を比較して単回帰分析を行う事で、予測する事を考えた。しかし得られたツイート数が少なく、そのうえ相関係数が最大0.5352と低いため正確な値を出す事が難しい。予測値が実際の店内人数と比べて大きく異なる結果となった。

精度を上げるためにツイートを増やす手法も検討したが、店名を用いないで今居る事が読み取れるツイートの取得が困難であるため行わなかった。取得した今居るツイート数の値に重みを付ける事で精度を上げる事にした。用いた重みは、前の時刻のツイート数が残存効果として

影響を与えていると考えて、30分と1時間前の今居るツイート数に重みを付ける手法を用いた。その結果相関係数が大きくなり、予測店内人数が実際の店内人数に近づいた。また、現時刻よりも1時間後のツイートが重視される事が分かった。

さらに精度を上げるために、時間帯を区切ってそれぞれに残存効果を適用し、単回帰分析を行った。その結果15時から18時は現時刻のツイート数が重視されるが、それ以外の店内人数が多くなる時間帯においては1時間後のツイート数が重視され、18時以降に至っては現時刻のツイート数を全く用いない結果となった。

本研究から言える事は、途中に立ち寄る場所を選ぶための場所を選ぶ判断基準として混雑度の取得を行い、手法としてツイッターの今居る事が読み取れるツイート数を用いて混雑度の予測を行ったが、結果としては不十分であった。しかし、前の時刻のツイート数に重みがある、ピーク時で予測値が実効値を下回る、他にも法則がありそれらを用いる事で、今後ツイート数を用いて混雑度を予測する手法の足掛かりはできたと考えている。これを元に精度の向上を行い、混雑度を元に途中に立ち寄る場所の混雑予測を行い、途中に立ち寄る事を考慮した経路検索システムを完成させたいと考えている。

謝辞

本研究に際して、様々なご指導を頂きました服部峻助教を初めとして、服部研究室の皆様にご感謝を致します。また、実験に使ったツイートを投稿した皆様、店内人数を調査した店にも感謝致します。そして、本研究で用いた TwitterAPI を提供している Twitter 社と、それを Java 上で動かせるライブラリを提供した Twitter4j に感謝致します。

参考文献

- [1] YAHOO! JAPAN 路線情報, <http://transit.yahoo.co.jp/> (2016).
- [2] ジョルダン, 青春 18 きつぷ検索, <http://www.jorudan.co.jp/norikae/seishun18.html> (2015).
- [3] Twitter, <https://twitter.com> (2016).
- [4] Shun Hattori, “Granularity Analysis for Spatio-Temporal Web Sensors,” WASET International Science Index (International Scholarly and Scientific Research & Innovation), Vol.7, No.2, pp.700–708 (2013).
- [5] Shun Hattori, “Spatio-Temporal Web Sensors Using Web Queries vs. Documents,” Journal of Automation and Control Engineering (JOACE), Vol.1, No.3, pp.192–197 (2013).
- [6] 永澤 勇樹, 吉田 京平, 服部 峻, “モバイル端末における旅行記の理解支援のための行程抽出と地図化,” 電子情報通信学会 モバイルネットワークとアプリケーション研究会, 信学技報, Vol.114, No.31, MoNA2014-4, pp.19–24 (2014).
- [7] 中野 隆介, 沼尾 雅之, “無線 LAN アクセスポイントへの検索要求を利用した鉄道車内混雑度推定,” 第 4 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 Forum 2012, A10-1 (2012).
- [8] 荻原 崇, 諏訪 敬祐, “Wi-Fi パケットセンサとクラスター分析を用いた屋内での混雑度推定手法,” 情報処理学会第 77 回全国大会, 2W-03, pp.3-315–316 (2015).
- [9] 音波を使ったチェックインソリューション「Air Stamp」の提供開始 (2015/12/21). https://www.nttdocomo.co.jp/info/news_release/2014/03/04_00.html
- [10] 寺田 賢治, 山口 順一, “カラーカメラを用いた通行人数計測システム,” 電気学会論文誌 D, Vol.118, No.3, pp.322–328 (1998).
- [11] 庭川 誠, 恩田 寿和, “監視カメラ映像中の局所的な動き検出とイベント累積による時間的かつ空間的な混雑度調査,” 電気学会論文誌 D, Vol.124, No.10, pp.1060–1066 (2004).
- [12] 総務省 | 平成 25 年版 情報通信白書 | 主な情報通信機器の普及状況 (世帯) (2015/12/29). <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h25/html/nc243110.html>
- [13] This recycling bin is following you (2015/12/29). <http://qz.com/112873/this-recycling-bin-is-following-you/>

- [14] JR 東日本アプリ (2015/12/21). <http://www.jreast-app.jp/>
- [15] 大野 宏, “万引き犯の行動分析と検知に関する研究,” 電気通信大学大学院電気通信学研究科 人間コミュニケーション学専攻 博士 (工学) の学位申請論文 (2009). http://ir.lib.uec.ac.jp/infolib/user_contents/9000000352/9000000352.pdf
- [16] Swarm, <https://ja.swarmapp.com/> (2015).
- [17] Wikipedia, “日本の鉄道駅一覧,” <http://ja.wikipedia.org/wiki/日本の鉄道駅一覧> (2015/12/18).
- [18] iタウンページ, <http://itp.ne.jp/>
- [19] Wikipedia, “鷺別駅,” <http://ja.wikipedia.org/wiki/鷺別駅> (2015/12/18).
- [20] 田原琢士, 馬強, “Twitter ユーザの活動するローカル地域の発見,” 情報処理学会第 77 回全国大会, pp.557-558 (2015). <http://id.nii.ac.jp/1001/00143976/>
- [21] 前田 高志ニコラス, 吉田 光男, 鳥海 不二夫, 大橋 弘忠, “Twitter 位置情報・テキスト情報を用いた人の移動モデル構築と観光地推薦手法の提案,” 第 9 回 SIG-DOCMAS 研究会, http://www.ai.soc.i.kyoto-u.ac.jp/docmas/proc/docmas-jsai-201511/maeda_takashinicholas.pdf
- [22] Twitter4J, <http://twitter4j.org/ja/> (2015).
- [23] Instagram, <https://www.instagram.com/> (2016).