

乗換案内データを用いた未来の混雑予測の研究

Irregular Congestion Prediction, not Detection

坪内 孝太^{*1} 下坂 正倫^{*2} 小西 達也^{*3} 丸山 三喜也^{*1} 山下 達雄^{*1}
 Kota Tsubouchi Masamichi Shimosaka Tatsuya Konishi Mikiya Maruyama Tatsuo Yamashita

^{*1}ヤフー株式会社 ^{*2}東京工業大学 ^{*3}東京大学
 Yahoo Japan Corporation Tokyo Institute of Technology The University of Tokyo

Thanks to the recent popularity of GPS-enabled mobile phones, modeling people flow has provided successful results in irregularity detection. With large-scale crowded events, it is crucial that there be enough time to take countermeasures against the irregular congestion, i.e., irregularity prediction.

To tackle this problem, we focus on route search logs. This paper presents a brand new framework for city-scale event prediction, not detection. By our observation of data where the route search logs related to a future event are in most cases repeatable and accumulated in proportion as the event draws near, we are able to leverage the divergence between the above two properties to predict city-scale irregular events. We demonstrate through experiments using the transit app logs that our approach can successfully predict city-scale crowded events one week in advance. The proposed method are evaluated and the method performs well to predict big scale irregular congestion

1. 序論

近年のモバイルデバイスの普及は目覚ましく、大量に収集されるユーザデータはいわゆるビッグデータとなり、各研究領域において専門家がそれらを分析し、定性的・定量的な分析を行うことが可能になってきた。特に、以前は専用の端末を用いてしか取得できなかった位置情報は、今やユーザのコンテキストを把握するのに一般的に使われるまでになった [Phithakkitnukoon 10, Gao 15, Schilit 94, Abowd 99]。現在までの研究成果から、大量の人の流れをモデリングし、断続的に大量に取得される位置情報を分析することで、定常的な地域特性を浮き彫りにすることが可能になり、店舗の出店計画や広告推薦、不動産売買などにも応用できるようになってきた [Fan 14, Gao 15, Cheng 13]。しかし時々刻々と激しく変化する現実社会に対しては、定常的な地域特性を把握するだけでなく、即応的な人口動態を解析する研究が必要とされていると言える。例えば、地震やハリケーンなどの大規模災害が起きたとき、人の流れを可視化することで現状の把握に役立ち、人命救助や都市機能の回復などへの貢献が期待される [Haddawy 15, Salfinger 15, Ludwig 15, Sakaki 10, Fan 14]。また、混雑を避けて比較的人の少ない経路を提示することも、人々の生活の質の向上や、排気ガス削減などに貢献できる [Dubey 15]。このような問題への挑戦として、人口動態の「異常検知」というドメインで研究が行われている。しかし異常検知技術は本質的に異常が生じてからの対応となるため、事前に何らかの対策を講じることができないのが欠点である。毎年多くの巡礼者が訪れる聖地メッカとその周辺地域では、将棋倒しによる死者が毎年のように発生しており、2015年のメナー群衆事故で2000人以上の人が圧死した事故は記憶に新しい。異常予測技術は、このような人為的なイベントで人口が集中したとき、誘導員の人員配置最適化や、公共交通機関の臨時増発計画にも利用できる [Wang 13]。

しかし、ある地域の日常的な人口は定常的なパターンを抽出する既存研究でも予測可能である [Shimosaka 15] が、大きなイベントが起きた場合に変動する人口を予測することは、定

常パターン抽出に特化しているがゆえに難しい。一方、最新の研究である Fan らの手法 [Fan 15] では1時間先の異常を予測することに成功し先駆的な成果と言えるが、予測可能な範囲はイベントの1時間前と非常に限定的であり、事前に対策を打つには時間が短いという欠点がある。ではなぜ異常人口予測は難しいのだろうか。その理由は、通常の日とイベントのある日とのデータ量に絶対的な差があり、イベントに関する定常的なパターン抽出が困難であるためである。その結果として通常時、異常時の人口を共に精度よく予測できないために、既存研究では異常時の人口予測を諦め、リアルタイムでの人口と通常時の推定人口を比較し、異常「検知」を行っているのである。

そこで本研究では、人口もしくはGPS履歴そのものからモデル化を行う研究とは対照的に、位置に関する予定情報を用いて異常予測を達成する。特に本研究で我々が着目したのは、ユーザの路線検索結果である。注目すべきは、異常を引き起こすイベントそれ自体は地域にとって「異常」な出来事でも、そこに向かう人はあくまで個人の予定に従って行動しているだけであるということである。そしてまた言うまでもなく、ユーザはイベントのない時でも自身の予定に従って移動行動を取っている。つまりユーザの予定情報を考慮することで、イベントの有無によらない統一的な枠組みの中で地域の人口モデリングを行うことが可能になる。一方、地域における乗換案内履歴の増減はイベントの有無により変化するものであり、曜日や天気などの要素によって規定される定常的なパターンが観測されることもまた事実である。異常検知では定常時の人口と実際の人口の差異に注目したが、定常時の人口と、実際の人口と強い相関性のある予定情報の差異は、いわば異常予測として機能する。本研究の貢献は以下の通りである。

- 位置履歴ではなく、位置に関する予定情報を用いることによる地域の人口予測モデリングを提案する世界初の研究である。
- 上記を応用することで、既存手法ではイベントの1時間前における予測が精々であった [Fan 15] が、異常予測を数日前の時点から可能であることを実験的に示した。

連絡先: 坪内 孝太, ヤフー株式会社, ktsubouc@yahoo-corp.jp

2. 関連研究

位置情報を用いた異常検出としては、例えば交通渋滞の検出手法についての詳細なサーベイがある [Dubey 15]. 渋滞をリアルタイム検出することでユーザーに迂回路を提示するというアプリケーションが考えられるが、渋滞の発生を「予測」し渋滞を未然に防ぐことは、これよりも更に抜本的な解決になることは言うまでもない. Isaacman ら [Isaacman 12] は、各ユーザー毎によく通る地域・経路を確率モデル化して考えることにより、都市レベルでの人口推定を可能にした. しかし、このモデルでは各ユーザーが同じ経路を何度も繰り返し通っていると仮定した訓練データを用いているため、普段行き着かない地域の人口推定には利用できないという問題点がある. Song ら [Song 13] は、東日本大震災のデータを用いて、災害時の人の避難経路推定手法を提案している. しかし、推定モデルに使われるデータは実際に災害が起きた後に収集されるものであり、災害発生時における推定には用い得ない. また、以前に起きた災害時のデータを使い回すことができないのも今ひとつ実用に及ばない. Sakaki ら [Sakaki 10] は、Twitter^{*1} のデータを使ってリアルタイム (1 秒以内) に地震発生を検知し、震源位置を推定するシステムを構築した. SNS 上の人々の活動がリアルワールドをよく反映していることを示す好例であり、避難経路などの予測技術が発展すれば、さらなる応用の可能性がある.

以上のように位置情報を用いた異常検出に関する研究は多いが、位置情報に限らずとも異常予測は難しい. state-of-the-art な研究として Fan ら [Fan 15] は、大きなイベントに開催に関して人の動きが通常時とは大きく異なることに着目し、1 時間の間の人の動きから次の 1 時間の人の流れを予測した. しかし予測を行えるのは 1 時間前という制約があり、都市レベルで異常となるような大きな変化に対して事前に対策を打てる程前に予測できないのが大きな問題である. 既存研究では位置情報に着目することで異常検出・予測を行ってきたが、位置に関する予定情報を利用することで、位置情報に基づくより高精度で長期的な異常予測が可能になると考えられる.

3. ユーザの予定情報を考慮した異常予知

3.1 乗換案内アプリのログデータ

ユーザーの予定情報を考慮した人口予測モデリングを行うために、本論文では Yahoo! JAPAN 経路検索のログデータを用いた. このログデータには、検索された経路上の駅をそれぞれ通過する時間帯が記録されており、これらを統合することで、いつどの駅にどれだけ人が滞在 (通過) するかが定量的に把握される. サービスの一番の特徴は過去・未来の任意の日時を指定できる点である. この未来の日時の検索ログに着目すれば、それはユーザーの「予定情報」と見なすことができる. 1 に示した、8/14 について 7/31,8/7,8/13,8/14 の時点で検索したクエリ数にはスケールに関して明らかに相関性がある. 本論文では、このように蓄積された予定情報に着目してモデルを構築する. ここで $x_{d,t|d-i}$ を、 $d-i$ 日の時点で d 日の時刻 t に関して検索した結果と定義する (すなわち、8/14 の 8 時について 7 日前に検索した結果は $x_{0814,08:00|0807}$). また、この手法における ground truth は検索日当日に関する検索結果 (e.g. 花火大会当日に、その日の何時の電車に乗るかを検索した場合) とした. 駅を通過する人数はその地域の人口と厳密には等しくないが、相関はあると考えられ、本研究の目的である異常予

測に関しては問題ないと考える.

3.2 提案手法の概要

既存研究で行われていた、異常「検知」は以下のような手順で行われていた. 1. 定常パターン抽出技術などを用いて地域の定常人口を予測する. 2. 実際の人口をリアルタイムもしくは事後的に取得する. 3. 上記 2 つの予測人口を比較し、それが乖離していれば異常と「検知」する. 地域における人口の異常を「検知」ではなく「予測」するためには、2. の部分を「実際の人口」から「異常時・定常時を含む短期的未来の人口 (予測値)」に置き換える必要がある. 本論文ではこの部分に、ユーザーの予定情報を考慮した予測モデルを用い、異常「予測」を試みる. 既存の定常的な人口を予測するモデルを DP (Descriptor-based Population model), ユーザの予定情報を考慮したモデルを SP (Schedule-based Population model) とする. DP は定常的な人口を予測するので、繰り返し頻度が低い異常 (赤線) に対して適切な予測を行うことができない.

一方で提案手法である SP モデルは 1 に示したように、過去時点での未来に対する検索数 (予定情報) と当日に行った検索数との関係をモデル化することで、異常人口に対して敏感な予測が可能になる. この SP モデルを利用した異常「予測」の具体的な手順は以下である. 1. は上と同じである. 2. ユーザの予定情報を考慮して、短期的な人口予測を行う. 3. 上記 2 つの予測人口を比較し、それが乖離していれば異常と「予測」する. 以下では、本研究で用いる 2 つの人口予測モデルについて述べる.

3.3 提案モデル

3.3.1 乗換案内のログ解析による定量的な異常予知

本研究では、上述の通り、定常人口予測モデル (existing) と、予定情報を考慮して定常時と異常時を統一的に人口予測するモデル (proposed) の 2 つのモデルを構築する. DP については Urban dynamics 分野で多くの研究がある. SP については他分野研究でも多く採用されている自己回帰的な考え方に基づくモデルである.

予測対象地域のある日、ある時刻の予測人口が「異常」に値するかどうかを表す定量的な指標として、その地域の通常時の人口と、予測された人口がどの程度乖離しているかを表す指標を定義する必要がある. 本論文では、日付 d と時刻 t に対して乖離率 $\nu_{d,t} = \frac{|\hat{x}_{d,t|d}^{SP} - \hat{x}_{d,t|d}^{DP}|}{\hat{x}_{d,t|d}^{DP}}$ を定義し、この乖離率と規定の閾値 $\bar{\nu}$ を用いて「異常」と判定することとした. このとき時刻 t は一定の時間間隔 Δ で離散化された時刻を表し、その値は経験的に調整する. ここで、 $\hat{x}_{d,t|d}^{SP}$, $\hat{x}_{d,t|d}^{DP}$ はそれぞれ SP, DP による日付 d の時刻 t に対する予測値である. ただし $\hat{x}_{d,t|d}^{DP} = 0$ のときは計算しない.

3.3.2 SP (Schedule-based Population) モデルによる将来の検索数の推測

1 節で述べたように、路線検索結果に含まれる「ユーザーの予定情報」を考慮することで通常時、異常時ともに精度よく地域の人口を予測する手法について論じる. ユーザの予定情報に着目した理由は、イベント開催に伴う人口の変動は地域としてみれば通常時と比べて「異常」でも、そのイベント参加者にとっては予定内の行動である、という仮定に基づくためである. 未来の日時に対する、数日前の時点でのユーザーの検索結果と、その教師データとなる当日の検索結果との相関関係をモデリングするというアプローチを取る. 本論文では、その簡便な手法として AR モデル [Akaike 69] を参考に、以下のような定式化を

*1 <https://twitter.com/>

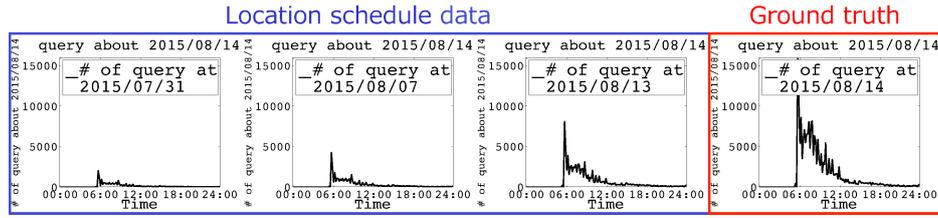


図 1: 蓄積した予定情報を利用した 8/14 の国際展示場駅の検索量予測

行う。

$$x_{d,t|d} = \sum_{i=1}^p \sum_{j=-q}^q \varphi_{i,j} x_{d,t-j|d-i} \quad (1)$$

すなわち、 q は時刻 t の予測に影響を及ぼす時刻の範囲を示す (i.e. $d-i$ 日時点で $t-q \leq t' \leq t+q$ なる時刻 t' に関して検索した結果が時刻 s の自己回帰に寄与するというのである)。また、推定対象である d 日当日に関して検索した結果は、上記と同様に $x_{d,t|d}$ などと記述できる。本論文においては AR モデルを用いて「予定情報と当日の検索数との相関関係をモデリング」している点が、通常の AR モデルを用いた人口予測の既存研究とは異なる SP のアプローチである。各パラメータは、負の対数尤度を最大化するように最適化を行う。

3.3.3 DP (Descriptor-based Population) モデルを用いた定常状態での検索量の推定

提案モデルである、ユーザーの予定情報を考慮した SP でどれだけ高精度に未来の人口が予測できたとしても、それが果たして地域にとって「異常」であるかどうかを判断、すなわち異常予測することはできない。これを行うためには、その地域の規範時の人口を予測し、それと比較することで異常判断をしなければならないためである。地域の規範時の人口予測に関しては、都市動態に関する既存研究で提案されているモデルを利用する。規範時の人口予測には、天気や曜日、祝日か否か、住民や通勤してくる人の属性など、様々な特徴量を用いて、地域の動態パターンを予測する。本論文では、state of the art な手法として、双線形低ランクポアソン回帰モデルを用いた都市動態推定手法 [Shimosaka 15] を利用する。このモデルは人口分布をポアソン分布で仮定し、天気や曜日などの外的要因 f と時間的要因 g を切り分けて双線形で特徴量化することで、都市動態パターンを表すポアソン分布のパラメータ $\lambda_{f,g} = \exp(f^T UV^T g)$ を表現する柔軟なモデルである。このとき、人口が $x_{d,t|d}$ である確率は以下のように表される。

$$p(x_{d,t|d}) = \text{Pois}(x_{d,t|d} | \lambda_{f,g}) = \frac{\lambda_{f,g}^{x_{d,t|d}} \exp(-\lambda_{f,g})}{(x_{d,t|d})!} \quad (2)$$

4. 実験

4.1 データの詳細

本論文の性能評価実験で用いるデータは、上述の通り Yahoo! JAPAN の経路検索サービスのログデータである。予測対象の駅を乗降者数などを参考に 103 駅に設定した。選定した駅の一例を表 1 に示す。

対象となる 103 駅に対してのみ、予測対象日の前日までのログを用いて、予測対象日の異常スコアを計算した。用いたデータの期間は 2015/11/3~2016/1/19 で、その間の乗換案内における総検索数は約 1 億件に達している。予測対象期間を

表 1: 選定した駅の例

新宿	品川	なんば	高槻	ドーム前
池袋	東京	国際展示場	市民広場	大阪城公園
渋谷	秋葉原	原宿	幕張本郷	有楽町
横浜	上野	銀座	名古屋	後楽園

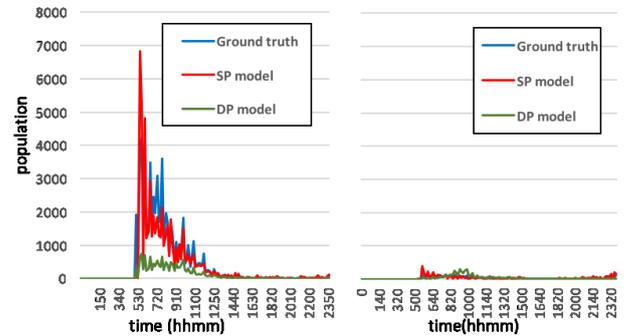


図 2: 検索数の一週間前からの予測 (左: コミケ当日/異常, 右: コミケ前日/平常)

2015/12/29~2016/1/19 とし、2015/11/3 から予測対象日の前日までを学習に使い、モデルを作成した。

また乗換案内サービスの特性として、ユーザーが目的地までの経路だけを調べるケースも考えられる。そのような場合、ユーザーは時間を気にしていないので現在時刻をインプットとして検索する。そのようなデータは「予定情報」とは言いえないため、日時指定をして検索したクエリのみ利用した。

4.2 SP モデルによる近い将来の検索量推定の検証

本研究の主目的である異常予測の評価の前に、ユーザーの短期的な予定情報を考慮することで、イベントのある日もない日も精度よく人口予測が可能かどうかを検証する実験を行った。

4.2.1 評価の結果

イベントにより生じた異常人口に関して特徴的な結果を取り上げる。2016 年 12 月に開催されたイベントであるコミックマーケット 91*2 は、3 日間で 55 万人が来場した大規模なイベントである。まさに異常混雑を、最寄り駅である国際展示場駅において上記 2 つの手法でどれほど予測精度に差が出るのかを観察するため、イベント最終日とその 1 週間後について予測した結果を 2 にプロットした。予測した異常に対して対策を打つというユースケースを想定し、1 週間前時点での予測を行った。右図 (定常時) を見てみると、DP と SP の予測に差は見られない。一方左図 (異常時) は約 100 倍以上のスケールだが、SP は精度良く人口予測できているが、DP では人口の増

*2 <http://www.comiket.co.jp/info-a/C91/C91info.html>

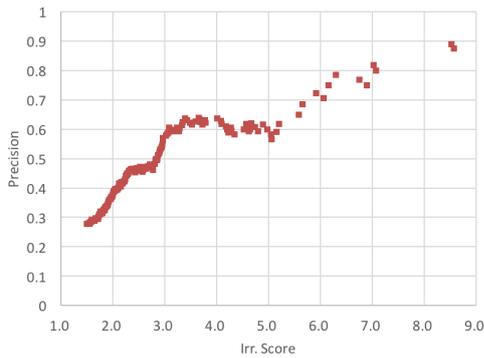


図 3: 異常予知スコアと予測精度との関係

加に対応できていない。またコミケのような大規模なイベントのない前日の 12/28 (右図) においては、既存研究である DP に加えて、提案手法である SP も精度良く人口予測できていることが確認できる。異常時には DP と SP の予測結果に差が生じているという実験結果は、DP と SP を比較することで異常予測が可能になるという我々の仮定を裏付けるものである。

4.3 性能評価

本手法の性能評価について図 3 に示す。図の横軸は Irr. Score (閾値 $\bar{\nu}$ と同義) を、縦軸はしきい値を Irr. Score 以上を異常と判断した際の、実際に異常混雑が起るイベントが発生していたのかを目視確認した際の精度である。例えば、 $\bar{\nu}$ が 7.0 以上のケースのみを異常と予測した際に、その精度は 0.8 であったという事を意味している。なお、目視確認は 3 名により行い、異常と判定された時間帯、駅に何か通常より人が集まりそうなイベントが本当にあるかをウェブやツイッター情報から調べた。

まず、図より、しきい値を高くすればするほど、推定精度は高くなっていることを示している。ところが、しきい値が 5.0 以上 (つまり、DP モデルと SP モデルの差が 5 倍以上) の場合でもその precision は 0.6 程度であり、しきい値が中程度および低い場合に設定した場合には、その精度は低くなる。たとえば、精度が 7 割程度のパフォーマンスが欲しいというケースでは、しきい値を 5.5 以上に設定する必要がある。つまり、平時よりも 5.5 倍以上の人が集まるようなイベントのみが推定対象となる。

5. 結論

本論文では、ユーザの予定情報 (乗換案内サービスの検索履歴) に着目することで短期的な人口予測を可能にする手法 (SP) を提案し、イベントの有無によらず人口予測が可能であることを実験的に検証した。次に、SP と定常人口 DP とを比較することで、今まで為し得なかった 1 週間前の時点での異常予測が可能であることを示した。本研究を応用して、ユーザに数日前の時点で迂回路を提示するなど「異常緩和」を行うこともできるようになる他、ユーザの「予定情報」を逐次的に学習することにより異常緩和策の効果測定も可能という点で画期的である。本研究では、ユーザの予定情報を利用して短期的な人口予測を行う SP と、都市の定常的な人口を予測する DP という 2 つのモデルを用いた。これらのモデルそれぞれの精度向上は結果として異常予測の精度にも繋がるものであり、今後の更なる研究が望まれる。本研究では非常に簡単な式で「乖離率」を定義することで定量的な評価を行ったが、この判定基

準をさらに洗練することで粒度の異なる異常予測が可能になると期待され、将来課題として考えられる。

また、大規模な異常混雑につながるようなイベントでは高い精度で検出できるが、異常の度合いが中規模 (平常時の 5 倍程度) のイベントになると急に精度が低くなることも確認した。このようなケースにも対応できる工夫も将来課題の 1 つといえる。

参考文献

- [Abowd 99] Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., and Steggles, P.: Towards a better understanding of context and context-awareness, in *HUC*, pp. 304–307 (1999)
- [Akaike 69] Akaike, H.: Fitting autoregressive models for prediction, *AISM*, pp. 243–247 (1969)
- [Cheng 13] Cheng, C., Yang, H., Lyu, M. R., and King, I.: Where You Like to Go Next: Successive Point-of-Interest Recommendation., in *Proc. of IJCAI* (2013)
- [Dubey 15] Dubey, P. P. and Borkar, P.: Review on techniques for traffic jam detection and congestion avoidance, in *Proc. of ICECS*, pp. 434–440 (2015)
- [Fan 14] Fan, Z., Song, X., and Shibasaki, R.: CitySpectrum: A non-negative tensor factorization approach, in *Proc. of UbiComp*, pp. 213–223 (2014)
- [Fan 15] Fan, Z., Song, X., and Shibasaki, R.: CityMomentum : An Online Approach for Crowd Behavior Prediction at a Citywide Level, in *Proc. of UbiComp*, pp. 559–569 (2015)
- [Gao 15] Gao, H., Tang, J., Hu, X., and Liu, H.: Content-Aware Point of Interest Recommendation on Location-Based Social Networks., in *AAAI*, pp. 1721–1727 (2015)
- [Haddawy 15] Haddawy, P., De Felice, G., Frommberger, L., Chakraborty, P., Saengpao, S., Kanchanakitkul, P., and Kauppinen, T.: Situation awareness in crowdsensing for disease surveillance in crisis situations, in *Proc. of ICTD* (2015)
- [Isaacman 12] Isaacman, S., Becker, R., Cáceres, R., Martonosi, M., Rowland, J., Varshavsky, A., and Willinger, W.: Human mobility modeling at metropolitan scales, in *Proc. of MobiSys*, pp. 239–252 (2012)
- [Ludwig 15] Ludwig, T., Reuter, C., Siebigteroth, T., and Pipek, V.: Crowdmonitor: mobile crowd sensing for assessing physical and digital activities of citizens during emergencies, in *Proc. of CHI* (2015)
- [Phithakkitnukoon 10] Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., and Ratti, C.: Activity-aware map: Identifying human daily activity pattern using mobile phone data, in *HBU*, pp. 14–25 (2010)
- [Sakaki 10] Sakaki, T., Okazaki, M., and Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors, in *Proc. of WWW*, pp. 851–860 (2010)
- [Salfinger 15] Salfinger, A., Girtelschmid, S., Proll, B., Retschitzegger, W., and Schwinger, W.: Crowd-Sensing Meets Situation Awareness: A Research Roadmap for Crisis Management, in *Proc. of HICSS*, pp. 153–162 (2015)
- [Schilit 94] Schilit, B., Adams, N., and Want, R.: Context-aware computing applications, in *Proc. of HotMobile*, pp. 85–90 (1994)
- [Shimosaka 15] Shimosaka, M., Maeda, K., Tsukiji, T., and Tsubouchi, K.: Forecasting Urban Dynamics with Mobility Logs by Bilinear Poisson Regression, in *Proc. of UbiComp*, pp. 535–546 (2015)
- [Song 13] Song, X., Zhang, Q., Sekimoto, Y., Horanont, T., Ueyama, S., and Shibasaki, R.: Modeling and probabilistic reasoning of population evacuation during large-scale disaster, in *Proc. of SIGKDD*, pp. 1231–1239 (2013)
- [Wang 13] Wang, L., Zhang, D., and Xiong, H.: effSense: energy-efficient and cost-effective data uploading in mobile crowdsensing, in *Proc. of UbiComp*, pp. 1075–1086 (2013)