犯罪データとSNSデータ可視化システムの提案

府川　和樹† 　中岡　佑輔† 　Panote　Siriaraya† 　王　元元†† 　河合由起子†
Adam Jatowt††† 　秋山　豊和†

† 京都産業大学コンピュータ理工学部　〒 603-8555 京都府京都市北区上賀茂本山
†† 山口大学大学院創成科学研究科　〒 755-8611 山口県宇部市常盤台 2-16-1
††† 京都大学 情報学研究科 社会情報学専攻　〒 606–8501 京都市左京区吉田本町
E-mail: †{g1445142,g1444936,kawai,akiyama}@cc.kyoto-su.ac.jp, ††spanote@gmail.com,
†††y.wang@yamaguchi-u.ac.jp, ††††adam@dl.kuis.kyoto-u.ac.jp

あらまし　本研究では，海外などの旅先で安全な経路推薦を目指し，発生した事件事故や犯罪に基づきツイートデータを場所と時間ごとに分析することで，危険な場所や通りの可視化および目的地までの安全な案内支援の実現を目的とする．提案する可視化システムでは，まず官公庁等が公開している犯罪データを取得し，次に犯罪発生時刻と場所（緯度経度）を抽出し，発生場所から最短距離となる通りを検出し，距離が閾値以内であれば通りにマッピングする．さらに，各犯罪場所で一定の発生時間内かつ一定距離圏内のツイートを取得し，記述内容から特徴語を抽出しタグクラウド形式で提示する．本可視化システムにより，犯罪データとツイートデータから交差点間で発生した犯罪発生数およびツイートデータ発生数の割合に応じて，交差点間の通りのラベリングが実現できる．本稿では，提案するナビの実現に向けサンフランシスコ市の犯罪に関するオープンデータおよびジオタグ付きツイートデータを取得し，発生時間間隔や発生場所周辺範囲に基づき地図上に提供できる可視化システムを提案し Web システムを構築する．また，犯罪データとツイートの発生割合を容易に比較可能なグラフや発生時間帯や発生場所に基づき抽出した特徴語のタグクラウドに関して構築した Web システムを用いて紹介する．

キーワード　犯罪データ，SNS データ，可視化，Web システム

1. はじめに

ツイッターに代表される SNS サービスの普及により，最新のニュースだけでなく身近な出来事など多種多様な大量の情報が発信されるようになった．それら発信されるメッセージは短いがリアルタイム性が高く，多くの人々によって発信された膨大な情報は，様々な分野の研究者や開発者にとって益々価値が高まってきている．Harshawardhan ら [1] がツイッターのデータ分析よりインフルエンザの傾向を追跡し予測しており，Healthcare の分野での利用価値は高まっている [2]．安全という側面では，同様にツイッターのデータ分析より地震との関連性の可視化が実践されている [3].

犯罪防止という側面でも SNS データの利活用が注目されている．犯罪防止の重要な一要素は，各犯罪行為に対する発生可能性の予測である．これまでの犯罪予測のアプローチでは，過去の犯罪データのみを分析し，発生割合の多い場所を犯罪が発生する可能性のある「犯罪スポット」として検出してきたが [4]，最近の調査ではツイッターのデータが犯罪予測に有用であることが分かってきた [5], [6]．つまり，犯罪を予測し防ぐのに Twitter の情報は一定の価値があると言え，特に犯罪現場や状況の調査を支援する上では，犯罪事件の特徴について貴重な情報提供となる可能性が高く，犯罪発生後も同様に有益な情報であると言える．また，専門家だけでなく知らない土地に訪問するユーザにとっても窃盗や発砲といった犯罪の種類に応じた犯罪データとそれに連動するツイート情報は重要な支援となり得る．

本研究では，危険な場所や通りの可視化および，目的地までの安全な案内支援の実現を目指し，発生した事件事故や犯罪に基づきツイートデータを場所と時間ごとに分析し可視化提供可能な Web システムの構築を目的とする．提案手法では，まず官公庁等が公開している犯罪データを取得し，それら犯罪発生時刻と場所（緯度経度）を抽出し，発生場所から最短距離となる通りを検出し，距離が閾値 dm 以内（d=20m）であれば通りにマッピングする．次に，各犯罪場所で発生時間前後 t 時間でかつ一定距離圏内のツイートを取得し，記述内容から特徴語を抽出する．抽出した特徴語より犯罪との関連性を分析する．さらに，犯罪データとツイートデータから交差点間で発生した犯罪発生数およびツイートデータ発生数の割合に応じて，交差点間の通りをラベリングする．本稿では，提案する通りの可視化および安全なナビの実現に向け，サンフランシスコ市の犯罪に関するオープンデータおよびジオタグ付きツイートデータを取得し，発生時間間隔や発生場所周辺範囲に基づいたマップによる可視化システムを提案し，構築した Web システムを紹介する．また，Web システムにおける発生時間帯や発生場所に基づき抽出した特徴語のタグクラウドや犯罪データとツイートの発生割合を容易に比較可能なグラフについても紹介する．

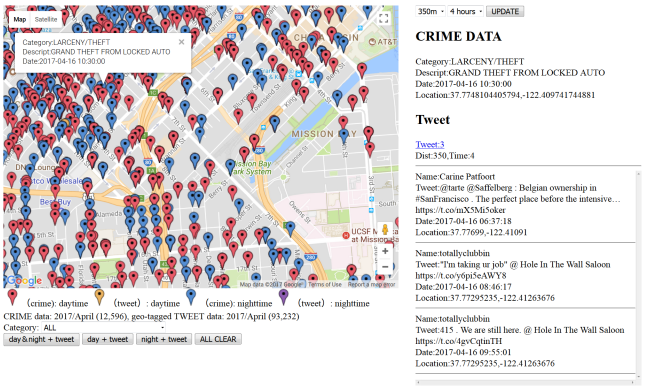

図 1　犯罪データと時空間で関連するツイート情報の可視化マップ

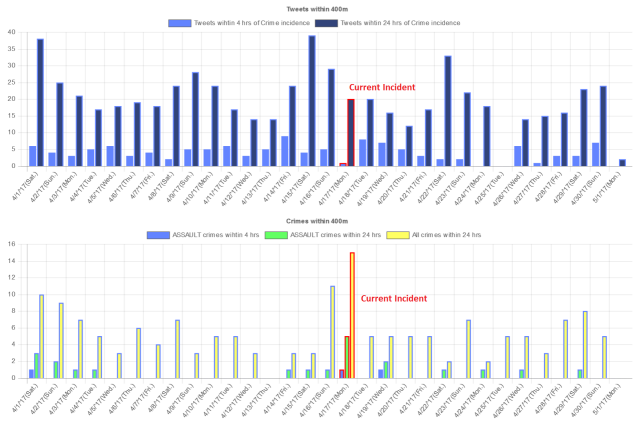

図 2　犯罪発生場所に対する犯罪発生総数とツイート発信総数の日別グラフ：（上）犯罪発生場所半径 dm 以内のツイート発信総数．（下）全ての種類の犯罪発生総数と同種類の犯罪発生総数

2.　犯罪・SNS データ可視化システム

本研究では，発生した事件事故や犯罪に基づきツイートデータを場所と時間ごとに分析することで，危険な場所や通りの可視化および，目的地までの安全な案内支援の実現を目的としている．提案する通りの可視化および安全なナビの実現に向け，犯罪データとツイートデータから犯罪発生時間間隔や発生場所周辺範囲に基づいたマップによる可視化システムを構築した(注1)．本章では，構築した Web システムを紹介する．

今回構築した可視化システムでは，米国のサンフランシスコの過去 14 年分（2003 年 1 月 1 日から 2017 年 6 月 25 日）の犯罪事件に関するデータ「Police Department Incidents Data Set(注2)」を取得し，そのうち 2017 年 4 月の 1ヶ月分を対象とした．なお，39 種類の違法犯罪となる犯罪データは 2,086,841 件であった．

2.1　犯罪データに基づく SNS データ抽出

構築した可視化システムのマップを図 1 に示す．マップの下の描画ボタンでは，39 種類の「犯罪の種類」が選択できる．犯罪の種類を選択した後，各犯罪発生時刻のうち「昼と夜」に発生した犯罪，「昼のみ」「夜のみ」に発生した犯罪および発信されたツイートを選択できる．なお，昼は朝 7 時から夜 6 時まで，夜は夜 6 時から朝 7 時までの時間帯とした．

図では，犯罪種別は「全ての犯罪」で「昼と夜」に発生した犯罪とツイートを描画した結果を示している．赤いピンは昼間に発生した犯罪でかつ時空間の関連ツイートがあった犯罪データを示しており，青色は夜に発生した犯罪でかつ時空間の関連ツイートがあった犯罪データを示している．時空間の関連ツイートとは，犯罪が発生した場所から半径 200m 圏内かつ発生時刻の前後 2 時間以内に発信されたツイートとした．なお，半径と時間幅は右上で設定を変更でき，半径は 50m から 400m の 50m おきに変更でき，時間は前後 1〜4 時間の 1 時間間隔で選択できる．また，時空間関連ツイートが発生していない犯罪データのピンは小さくグレーで表示した．

ピンを選択しクリックすると，選択された犯罪データが地図の右側に「CRIME DATA」として表示される．また，同時に時空間関連ツイートが「Tweet」として表示される．図では 3 件のツイートが提示されている．

マップにより，例えば，昼間に犯罪の多い地域や通りを効率的に把握でき，専門家だけでなく旅行者にとっても有益な情報支援につながると考えられる．

2.2　犯罪発生場所・時間に基づくグラフ生成

前節より，犯罪データと時空間で関連するツイートのマップへの可視化を実装し，犯罪データによる危険な通りや場所の情報提供を実現した．次に，犯罪とツイートとの発生の関係性の理解支援を目的に，犯罪発生場所における犯罪発生数とツイート発生数をグラフとして可視化した．

図 2 にマップ上のピン（犯罪データ）をクリックした際にリアルタイム生成されるグラフの例を示す．グラフは横軸が日付で縦軸が総数となる．上のグラフはツイート発信総数で下のグラフは犯罪発生総数である．また，両グラフとも赤で囲まれた日がクリックした犯罪発生日を示す．

ツイート発信総数のグラフ（上）の濃青は選択した犯罪発生場所から半径 dm 以内で同日 24 時間に発信されたツイート総数で，薄青は半径 dm 以内の前後 2 時間以内に発信されたツイート総数を示す．犯罪発生総数のグラフ（下）の黄色は犯罪発生場所から半径 dm 以内で同日 24 時間以内に発生した全種類の犯罪総数で，緑は 24 時間以内でかつ同種類の犯罪のみを示し，青は同種類の犯罪のみでかつ前後 2 時間以内の犯罪総数を示す．

今後，得られた犯罪発生数とツイート発信数との相関を犯罪の種類やツイートの発信時刻に対して検証する予定である．

2.3　時間と場所に基づくツイートの特徴語抽出

前節では犯罪発生数とツイート発信数の可視化を実装した．本節では，発生した犯罪に対する時空間関連ツイートの特徴語を抽出する．特徴語は，出現頻度（TF-IDF）より算出した値を用いて下記の 3 種類のタグクラウド形式で可視化した．

- 犯罪の時間的関連となるツイート特徴語
- 犯罪の近辺の空間的関連となるツイート特徴語

(注1)：http://delab.kyoto-su.ac.jp/~yasui/map/CrimeMap.html
(注2)：https://catalog.data.gov/dataset/sfpd-incidents-from-1-january-2014

図 3　発生した犯罪の時空間関連ツイートより生成した 3 種類のタグクラウド

- 犯罪の時空間的関連となるツイート特徴語

図 3 にマップ上のピン（犯罪データ）をクリックした際に生成されるタグクラウドの例を示す．

左上のタグクラウドは，発生した犯罪の時間的関連となるツイートの特徴語の抽出結果である．算出法は，特定の時間範囲（例えば前後 2 時間）以内送信されたツイートに使用された単語の頻度（TF）と，同じ場所（例えば犯罪発生 200m 圏内）で同じ時間範囲以内の全日数（IDF）とした．

右上のタグクラウドは，犯罪の発生場所の近辺の空間的関連となるツイートの特徴語抽出結果である．算出法は，犯行が発生した場所半径 dm 圏内で特定の時間範囲内から送信されたツイートに使用されている単語の頻度（TF）と，同じ時間帯で同日に全領域で発信されたツイート（IDF）とした．

下のタグクラウドは，犯罪の発生場所と時間と関連するツイートの特徴語抽出結果である．算出法は，犯行が発生した場所半径 dm 圏内で特定の時間範囲内から送信されたツイートに使用されている単語の頻度（TF）と，同じ時間帯で全日数でかつ全領域で発信されたツイート（IDF）とした．

今後，抽出した特徴語と発生した犯罪内容との関連性を抽出することで，時空間的関連だけでなく内容との関連性の高いツイート検出および可視化を目指す．

3. ま と め

本研究では，専門家への犯罪分析のための情報支援および目的地までの安全な案内支援の実現を目的とし，発生した事件事故や犯罪に基づきツイートデータを場所と時間ごとに分析することで，危険な場所や通りの可視化システムを提案構築した．本稿では特に，犯罪データとツイートデータから犯罪発生時間間隔や発生場所周辺範囲に基づいたマップによる可視化システムを紹介した．構築したシステムを利用することで，時間や場所ごとに犯罪の多い地域や通りを効率的に把握できることを確認できた．今後，得られた犯罪発生数とツイート発信数の相関を犯罪の種類やツイートの発信時刻に対して検証する予定である．また，犯罪データとツイートデータの時空間的関連だけでなく，抽出した特徴語と発生した犯罪内容との関連性検出を行う予定である．

謝　　辞

本研究の一部は，総務省 SCOPE（受付番号 171507010），JSPS 科研費 16H01722，17K12686 の助成を受けたものである．ここに記して謝意を表す．

文　　献

[1] Harshavardhan Achrekar，Avinash Gandhe, Ross Lazarus, Ssu-Hsin Yu and Benyuan Liu,“Predicting flu trends using twitter data”，IEEE pp. 702-707，2011.

[2] Abeed Sarker, Karen O ’ Connor, Rachel Ginn, Matthew Scotch , Karen Smith , Dan Malone and Graciela Gonzalez ,“Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter.”，Springer,Drug safety（39-3），pp. 231-240，2016.

[3] Takeshi Sakaki , Makoto Okazaki and Yutaka Matsuo,“Tweet analysis for real-time event detection and earthquake reporting system development.”，IEEE,IEEE Tran. on Knowl. and Data Engin （25-4），pp. 919-931，2013.

[4] Spencer Chainey, Lisa Tompson and Sebastian Uhlig ,“The utility of hotspot mapping for predicting spatial patterns of crime.”，Springer,Security Journal（21-1-2），pp. 4-28，2008.

[5] Xiaofeng Wang , Matthew S Gerber and Donald E Brown,“Automatic Crime Prediction Using Events Extracted from Twitter Posts.”，Springer,SBP （12），pp. 231-238，2012.

[6] Matthew S Gerber,“Predicting crime using Twitter and kernel density estimation”，Elsevier, Decision Support Systems（61），pp. 115-125，2014.