

(30) 地域の安全知情報の獲得のための CGM解析手法に関する研究

山本 雄平¹・中村 健二²・田中 成典³・藤本 雄紀⁴

¹非会員 関西大学大学院 総合情報学研究科 (〒569-1095 大阪府高槻市霊仙寺町2-1-1)

E-mail: yamamoto@kansai-labo.co.jp

²正会員 大阪経済大学准教授 情報社会学部 (〒533-8533 大阪市東淀川区大隅2-2-8)

E-mail: k-nakamu@osaka-ue.ac.jp

³正会員 関西大学教授 総合情報学部 (〒569-1095 大阪府高槻市霊仙寺町2-1-1)

E-mail: tanaka@res.kutc.kansai-u.ac.jp

⁴非会員 関西大学 総合情報学部総合情報学科 (〒569-1095 大阪府高槻市霊仙寺町2-1-1)

E-mail: fujimoto@kansai-labo.co.jp

各都道府県警にて提供される犯罪情報マップは、街頭犯罪や迷惑行為から地域住民の安全を守るために、安全知を住民に提供している。しかし、犯罪情報マップで公開される情報は、現行犯や住民通報の情報であり、警察が認知していない情報は含まれていない。一方、CGMには、犯罪情報マップに含まれない様々な安全知が公開されており、住民に提供することで、安全・安心な街づくりに役立つと考えられる。そこで、本研究では、CGMを詳細に分析し、安全知を抽出する手法を提案する。さらに、各投稿に含まれる位置情報から街頭犯罪マップの構築を行い、CGMに含まれる安全知の提供の可能性を検証する。

Key Words : CGM, Web Page Segmentation, Street Crime Map, GIS

1. はじめに

街頭犯罪や迷惑行為などの増加により、地域住民の安全が脅かされている。これらの犯罪から地域住民の安全・安心を守るための取り組みとして、警視庁から提供されている事件事故発生状況マップ¹⁾や各都道府県警から提供されている犯罪発生マップ²⁾がある。これらのサービスは、地域の安全・安心なまちづくりのための安全知を提供しており、地域の防犯活動に利用されている。しかし、これらのサービスで公開されている情報は、実際に発生した事件や住民の通報により提供された情報であり、警察が認知していない情報は含まれていない。例えば、街頭犯罪や迷惑防止条例³⁾違反に関連する情報(公共の場におけるたむろ情報、盗撮スポットなど)などは含まれておらず、地域の安全・安心な環境が脅かされている状況である。

上述の街頭犯罪や迷惑防止条例違反に関連する情報は、CGM (Consumer Generated Media) と呼ばれるインターネット上のクチコミサイトや電子掲示板等の誰でも情報発信が可能なWebサイトで情報交換されるケースが多くみられる。そのため、これらの情報を解析することで地域

住民の安全が向上できると考えられる。このための手法として、CGMを解析する研究⁴⁾が行われている。これらの研究では、Webページに含まれる全ての内容を対象として情報の抽出や判定等を行っている。しかし、犯罪に関する情報は、投稿毎にその犯罪の発生した場所や時間、内容が異なるため、Webページに含まれる全ての内容を解析対象とする手法では解析が困難である。そのため、犯罪に関する情報をCGMから抽出するには、Webページを投稿単位に分割し、情報の解析を行うことで対応できると考えられる。そこで、本研究では、Webページを投稿単位に分割する手法をCGMに適用し、各投稿を解析する手法を提案する。Webページの分割手法を適用することにより、各投稿に含まれる位置情報を考慮して投稿を解析できるため、GISを用いて従来の犯罪発生マップなどにこれらの情報を掲載する事が可能である。そのため、本研究では、CGMから抽出した情報に基づいて街頭犯罪マップを構築し、本提案手法の有効性を検証する。

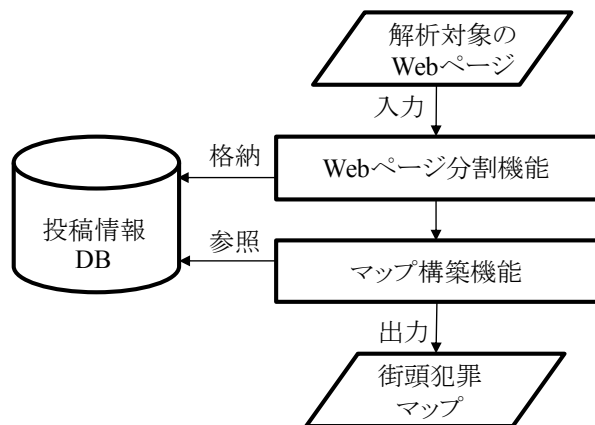


図-1 処理の流れ

2. 研究概要

(1) 研究目的

本研究では、CGMの投稿毎の情報に基づく街頭犯罪マップを構築する。本提案手法を用いることで、従来の犯罪発生マップでは公開されていない情報を分析する事が可能であると考えられる。本研究では、街頭犯罪マップを構築するために、Webページに含まれる各投稿の情報が必要となる。そこで、本研究では、Webページを投稿単位に分割する手法を利用して、CGMから各投稿を抽出する。そして、位置情報に基づいて各投稿をGIS上にマッピングし、街頭犯罪マップを構築することで、CGMに含まれる安全知が提供可能であるかを検証する。

(2) 処理の流れ

本研究の処理は図-1に示すとおり、Webページ分割機能とマップ構築機能の2つの機能で構成される。各機能の処理手順を次に示す。

Webページ分割機能は、図-1に示す通り入力された解析対象のWebページを投稿単位に分割し、投稿情報DBにWebページに含まれる各投稿を格納する。本研究で用いるWebページの分割手法は、様々なWebページに対応するため、DOM (Document Object Model) を用いる手法⁹⁾を採用する。本機能では、まず、入力されたWebページのDOMを構築する。構築したDOMを対象としてWebページ分割手法を適用し、Webページを投稿単位に分割する。最後に、分割した各投稿を投稿情報DBに格納する。

マップ構築機能は、Webページ分割機能で分割した各投稿情報を取得し、GIS上に投稿情報と共にマッピングする。本機能では、まず、位置情報を抽出するために、各投稿のテキストに対して形態素解析を行い、地域に関する形態素を位置情報として抽出する。ここで、抽出した地域に関する形態素が複数回連続して出現した場合は、形態素を結合したものを位置情報として利用する。次に、抽出した位置情報からジオコーディングAPIを利用して緯度経度を算出する。そして、算出した緯度経度に基づ

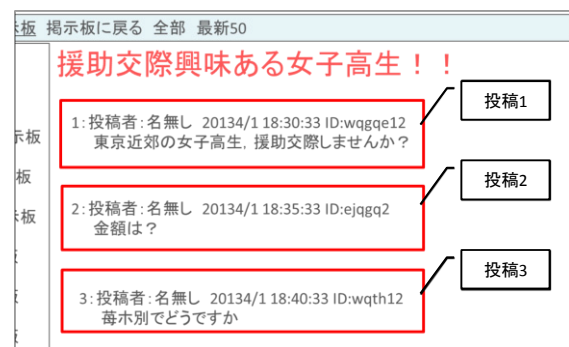


図-2 Webページ分割のイメージ

き、投稿情報と共にGIS上へマッピングし、街頭犯罪マップを構築する。

3. Webページ分割アルゴリズムの選定

本研究で用いるWebページ分割手法は、既存研究⁸⁾で提案されている手法を採用する。Webページの分割イメージを図-2に示す。既存研究⁸⁾で提案されているWebページ分割アルゴリズムは、DOMから算出したWebページの文字や画像等の要素を表示されている位置と面積の包含関係に基づきグループ化するアルゴリズムである。電子掲示板等のWebページを目視で確認する場合、そのWebページに表示されている内容から見た目により各投稿を判断する。そのため、本手法を用いることで、Webブラウザに表示される内容からWebページを分割できるため、Webページを投稿単位で分割して抽出可能であると考えられる。なお、Webページ分割アルゴリズムの詳細は、既存研究⁸⁾を参照されたい。

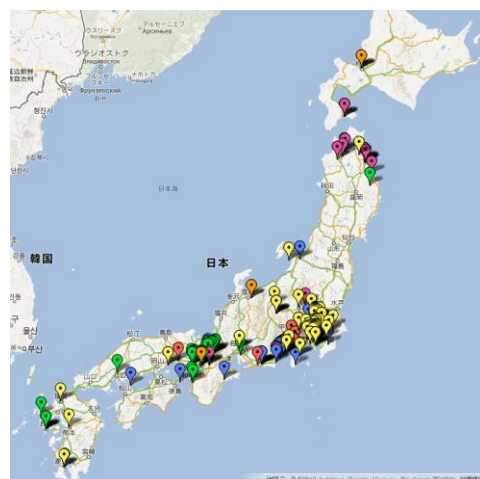
4. 街頭犯罪マップの試作

(1) 実験内容

本実験では、本手法の有用性を検証するために、投稿情報に基づいて街頭犯罪マップを構築し、その内容について分析する。街頭犯罪マップを構築する際の投稿情報は、実際のWebページを収集し、そのWebページを投稿単位に分割した情報を用いる。本実験では、Webページに含まれる全ての情報を利用して構築した街頭犯罪マップ（以下、Webページ全体手法）とWebページに含まれる投稿単位の情報を利用して構築した街頭犯罪マップ（以下、本提案手法）とを目視で確認し、それぞれの手法で構築した街頭犯罪マップを分析することで本提案手法の有用性を検証する。なお、本実験では、Webページ全体手法と本提案手法とで構築した街頭犯罪マップの分析を行うため、Webページ全体手法の街頭犯罪マップを構築する必要がある。そこで、本実験では、図-1のマッ



Webページ全体手法:13件



本提案手法:716件

図-3 本実験で構築した街頭犯罪マップ

ブ構築機能にWebページに含まれる全てのテキストを入力し、Webページ全体手法の街頭犯罪マップを構築する。

(2) 実験対象のデータ

本研究では、CGMを対象としているため、様々な投稿者が情報を投稿する電子掲示板を実験データとして用いる。本実験で対象とするWebページは、任意に選定したキーワードでGoogle掲示板検索⁹⁾を行った結果、犯罪に関する情報が投稿されている電子掲示板のWebページ13件を用いる。Webページを収集するために用いたキーワードは、街頭犯罪に関する情報が得られると考えられるキーワードを人手により選定する。

(3) 実験結果と考察

a) 実験結果の比較

本実験で構築した街頭犯罪マップを図-3に示す。図-3左はWebページ全体手法で構築した街頭犯罪マップであり、図-3右が本提案手法で構築した街頭犯罪マップである。また、本実験では、選定したキーワード毎に色分けして街頭犯罪マップを構築した。

図-3を確認すると、Webページ全体手法により構築した街頭犯罪マップと比較して、本提案手法で構築した街頭犯罪マップの方が、多くの情報をマップ上にマッピングできていることがわかる。それぞれの街頭犯罪マップにマッピングされた投稿情報の件数を確認すると、Webページ全体手法では対象としたWebページ件数と同じ13件であるに対して、本提案手法は716件となっており、記事単位に分割して詳細化することで約55倍の情報を提供できていることが明らかとなった。このことから、特に電子掲示板形式のWebページを解析し情報を取得する際には、本研究のWebページの分割手法が効果的なことがわかった。

b) 本提案手法で構築した街頭犯罪マップの分析

街頭犯罪や迷惑行為に関する情報を適切に抽出するた

表-1 Webページに含まれる情報の内訳

項目	件数	内訳	
有害又は安全知として必要	24件	地図上にある	7件
		地図上にない	17件
無害又は安全知として不要	27件	地図上にある	11件
		地図上にない	16件

めには、投稿内容を詳細に分析する必要があると考えられる。そこで、今後の発展研究に繋げるために、実験で用いたWebページの中から無作為に1ページ分のデータを抽出し、投稿単位に分割した結果得られた51件について目視で確認した。投稿内容に基づきそれぞれの投稿を人手で集計した結果を表-1に示す。集計結果を確認すると次に示す2つの内容が明らかとなった。

- GIS上に適切にマッピングするための工夫が必要なのことがわかった

表-1を確認すると、安全知として必要な情報は24件あり、7件の情報は地図上にマッピングできている事がわかった。一方、地図にマッピングできていない17件の情報を確認したところ、地域名やランドマークに関する情報が投稿情報の中に含まれていない例や、地域名やランドマークに関する情報の一部が伏字で表現されており位置情報の取得に失敗している例が見られた。これらの情報に対応するためには、前者の例には、投稿内容に関する他の情報を収集し、参照することにより投稿内容に合致する地域名を抽出する手法の考案が必要であると考えられる。後者の例には、伏字を含む位置情報に類似する地域名やランドマーク名を推定する事で対応できると考えられる。

正しく提供できた安全知情報の例を表-2に示す。これらの情報では、位置情報に関する内容が投稿に含まれており、投稿内容の中で話題となっている位置情報が限定されている事がわかる。このことから、マッピングする

表-2 正しく提供できた安全知情報の例

犯罪情報の種類	投稿内容
盗撮に関する内容	ミニスカ好きの彼女を梅田で連れまわしてたけど、俺がエスカ先に乗って彼女を後ろにのさせるとその彼女の後ろにピタリ張り着いてるやつは盗撮野郎か阪急んとこなんて実際にスマホ片手に彼女のスカート下にしてたから撮ってたな。
車上荒らしに関する内容	今日入谷駅の駐輪場に停めてた原付の車上荒らしにあったわ。
痴漢に関する内容	小中野とか青葉とか諏訪って変な人よく出没しない？痴漢や露出、パンツ売ってくれとか。だから暗くなったらあそこらへんは怖くて一人で歩けない。痴漢にあった時はしばらく人が怖かった。

上で欠けている情報を補足することで、さらに多くの情報がマッピング可能であると考えられる。

● 掲載する情報の選定が必要なことがわかった

表-1の「無害もしくは安全知として不要」の項目に含まれる情報を確認すると、安全知としての情報価値は無いが、位置情報が含まれているためマッピングされている投稿が 11 件あることがわかった。本研究で構築した街頭犯罪マップは、安全知に関する情報を提供する地図であることから、これらの情報が含まれている事により混乱を招く可能性がある。そのため、これらの不要な投稿情報は、地図上にマッピングする前に処理対象から除外することで、安全知として有用な情報のみを可視化した街頭犯罪マップを構築できると考えられる。

5. おわりに

本研究では、Web ページ分割手法を用いることで、Web ページ全体を対象としていた場合と比較して詳細な安全知を獲得し、地域住民に提供できる可能性を示唆した。これらの安全知に関する情報は警察機関やその他の関係機関が提供できない情報であり、安全・安心なまちづくりへの一助となると考えられる。その一方で、いくつかの課題も見られた。

1 つ目の課題として、不要な情報が地図上の多く含まれていた事が挙げられる。本課題に対しては、各投稿の内容が安全知として必要な情報であるかどうかを自動で判定する手法を検討する必要があると考えられる。安全知に関する情報は、有害な情報に限らず、ニュースに対する意見等、様々な情報が該当すると考えられる。そのため、情報の判定には、単純に有害かどうか以外の基準も検討する必要があると考えられる。

2 つ目の課題として、地図にマッピングできない情報がある事が挙げられる。これらの情報は、投稿内容に地域名やランドマークに関する情報が含まれていない事が主な原因であることがわかった。そのため、CGM がコミュニケーションツールであることから、各投稿間の関連を解析することにより、位置情報の取得が可能である

と考えられる。

今後は、上記の 2 つの課題を解消するとともに、情報の信頼性を評価する方法について検討する予定である。実際に安全知に関する情報は多いほど良いが、信憑性に欠ける情報は、危険を呼ぶ可能性がある。そのため、今後は、この点を考慮した街頭犯罪マップを構築する予定である。

参考文献

- 1) 警視庁：事件事故発生状況マップ，<<http://www.keishicho.metro.tokyo.jp/toukei/jikommap/jikommap.htm>>，(参照 2013.6.21).
- 2) 警視庁：犯罪情報マップ，<<http://www.keishicho.metro.tokyo.jp/toukei/johomap/johomap.htm>>，(参照 2013.6.21).
- 3) 大阪府：大阪府公衆に著しく迷惑をかける暴力的不良行為等の防止に関する条例，<http://www.pref.osaka.jp/houbun/reiki/reiki_honbun/ak20110671.html>，(入手 2013.6.21).
- 4) 中村健二，田中成典，大谷和史，山本雄平：セキュアライフの創出を目指した安全知の獲得に関する研究—電子掲示板からの犯行予告の抽出—，土木情報利用技術論文集，Vol.18，pp.269-280，2009.
- 5) Orkut, B., Hector, G.M. and Andreas P.: Seeing the Whole in Parts: Text Summarization for Web Browsing on Handheld Devices, Proceedings of the 10th International Conference on World Wide Web, ACM, pp.652-662, 2001.
- 6) 伊藤太樹，浅見昌平，大田忠親，新谷虎松：SVM に基づくテンプレートを考慮した Web ページの分割手法について，人工知能と知識処理研究会技術研究報告，電子情報通信学会，Vol.108，No.119，pp.81-86 2008.
- 7) Cai, D., Yu, S., Wen, J.R. and Ma, W.Y.: Extracting Content Structure for Web Pages Based on Visual Representation, Proceedings of the 5th Asia-Pacific Web Conference on Web Technologies and Applications, Springer-Verlag, pp.406-417 2003.
- 8) 中村健二，田中成典，山本雄平，安彦智史：共起関係の抽出範囲を考慮した有害情報フィルタリング手法，情報処理学会論文誌，情報処理学会，Vol.54，No.2，pp.571-584，2013.
- 9) Google Inc.: Google, <<http://www.google.co.jp/>>, (参照 2013.6.21).