安全・安心メールの自動分類と警察統計との比較による対策利用価値の考察

中央大学大学院 学生会員 ○齋藤 勝久
東京大学生産技術研究所 正会員 沼田 宗純
東京大学生産技術研究所 正会員 目黒 公郎

１．目的

効果的な防犯対策は犯罪情報全体を俯瞰・評価した上で立案されるべきであり,その一部としての防犯活動は対象地域の特性と過去に発生した犯罪情報の分析に基づいて実施することが大切である.そこで本研究では，地域の犯罪状況をリアルタイムに収集,蓄積し,迅速に分析できる環境の構築を目指す.具体的には現在全国各地で普及している安全・安心メールと呼ばれる行政サービスを利用し,地域のリーダーが独自に地域活動の分析や評価ができるシステムを構築した.

安全・安心メールについての厳密な定義はないが,本研究では「警察又は行政が配信する安全・安心に関する住民向けのメールサービス」,または「警察や行政が HP 上で呼びかける住民への危険についての詳細が書かれた情報」と定義した.安全・安心メールから犯罪統計を作れば,統計のデータ元を確認することができる.さらに安全・安心メールは短文ながら,「どこで,誰が,誰に対して,どのような行為が発生したのか」をほぼリアルタイムに知ることができるため,様々な種類の犯罪統計を毎日更新することが可能である.また,安全・安心メールには法で定義される犯罪に至らなかった情報も含まれるため,件数が多くなり,地域を限定しても統計的に有意な数の情報を集めやすい.

安全・安心メールの分類は齋藤らの研究[1)]を参考に,犯罪事象,時間,場所,被害者に対してそれぞれ行った.そして得られる統計データがどのような意味を持つかを考察した.

２．安全・安心メールの分類

全国 111 の配信団体が提供する安全・安心メールを対象に情報収集した結果,41,794 件のデータが収集できた.収集した情報に,事象,空間,被害者と言った属性を付与し分類した.大量のメールを自動処理するために,項目毎に分類を決定するキーワードを用意した.そして,キーワードが文章中に登場するか否かを判定し,該当した項目があればその属性を付与した.住所を除く分類の項目数は 355 個,それに関するキーワードは 1,522 個になった.犯罪事象を持たないメールを犯罪以外の安全・安心についてのメールと見なした結果,メールの件数は 24,568 件に絞られた.この中の 350 件のサンプルを任意に抽出し,その内容を検証した.結果を表 1 に示す.包括罪種については,専門家による分類を正解として,自動分類と比較した結果,正解率は 98%となった.他の分類も 79%以上の正解率となっている.ただし,分類によってはメールの記述情報が少ないために専門家も自動判別も共に判別不能となったケースも多い.

表 1 機械的な分類の精度

分類	項目例	正解率	判別不能情報率
包括罪種	粗暴犯	98%	0%
包括罪種-内訳罪名	暴行	98%	10%
犯罪累計-行為	殴る	97%	10%
施設	街頭	80%	41%
施設名称	公園	79%	43%
設備	トイレ	94%	91%
(被)性別	男	84%	52%
(被)年代	小学生	91%	63%
(被)年代詳細	小学5年生	99%	97%

３．安全・安心メールと警察統計との比較

メールの分類結果を平成 21 年上半期(2009 年 1 月 1 日から 6 月 30 日)の犯罪情勢警察統計[2)]と比較した.

1) 被害者の年代別件数

被害者の年代を比較した結果を表 2 に示す． 安全・安心メールは被害者が中学生以下の犯罪に関する情報が多いことがわかった.

2) 施設名称に対する比較

次に小学生と中学生が被害者となっている犯罪の

キーワード 犯罪発生情報,防犯活動，ボランティア活動，安全・安心メール，統計分析，子ども
連絡先 〒153-0041 東京都目黒区駒場 4-6-1 東京大学生産技術研究所 Be-604 Tel03-5452-6436

表 2　被害者年代別件数比較

年代(件)	メール	警察統計
未就学児	19	186
小学生	1416	11428
中学生	616	28215
その他少年	986	87374
社会人	0	539805
詳細不明	5411	0
計	8448	667008

みを対象に,安全・安心メールと警察統計の発生犯罪について,犯罪が発生した施設名称の件数を比較した(表 3).安全・安心メールと警察統計の統計データの関係を見るために,件数を対数変換した後,警察統計と安全・安心メールの相関係数を求めると,被害者が小学生の場合 0.52 ,中学生の場合 0.26 となり弱い相関があった.低い相関係数を示した中学生のデータに注目すると,駐車場について,安全・安心メールと警察統計の乖離があった.また,全般的に建物に関連する犯罪が,安全・安心メールからの情報だけでは不十分なことがわかった.

3) 罪種に対する比較

上と同様に,安全・安心メールと警察統計の発生犯罪について,罪名件数を比較した(表 4).安全・安心メールと警察統計の統計データの関係を見るために,上と同様に相互の相関係数を求めると,警察統計と安全・安心メールの被害者が小学生の場合 0.59 ,中学生の場合 0.65 となった.窃盗,傷害に関する項目も,警察統計に比べ安全・安心メールには含まれないことがわかった.一方,わいせつの項目は安全・安心メールに多く含まれることがわかった.

警察統計と安全・安心メールの差は,特に窃盗犯,駐車（輪）場に表れるが,これは安全・安心メールに自転車盗などが含まれる駐車場で発生する窃盗事件が反映されないことが原因と推測できる.また,警察統計に多く表れる暴行や傷害事件が安全・安心メールで少ない要因の一つは,安全・安心メールでは不審者情報が中心であるため,身内,知合いが加害者となる犯罪があまり含まれないことが挙げられる.全ての項目の中で,唯一警察統計より多くの被害件数があったわいせつに関しては,警察で取り扱わないような軽微なわいせつ事件が小中学生を対象として発生し,安全・安心メールがこれを配信しているためと推測できる.

表 3 警察統計と安全・安心メールの施設名称件数比較

被害者	小学生		中学生	
媒体(件)	警察統計	メール	警察統計	メール
一戸建て住宅	1024	35	2150	7
共同住宅	2001	16	3211	2
学校(幼稚園)	247	19	1167	4
駐車(輪)場	3756	30	12194	14
道路上	1910	869	4823	479
都市公園	1113	158	684	14
空き地	52	3	179	2
列車内	9	21	41	15
駅・鉄道施設	28	17	289	11
その他	1288	65	3477	14
計	**11428**	**1233**	**28215**	**562**

4. まとめ

安全・安心メールは短文であるため分類が容易で,大量の情報を集めやすいというメリットがある一方で,情報の中には分類に必要な最低限の被害状況が記述されていない場合もある.現状ではボランティアが統計を対策に生かそうとすると,わいせつ犯に関しては安全・安心メールからだけでも十分な情報を得ることができるが,他の犯罪,特に窃盗犯に関しては情報の公開が限定的である警察統計を参考にするしかないことがわかった.

5. 今後の課題

今後は警察統計と安全・安心メールの関係性についての分析をさらに進める.また,より多くの情報収得を目的として,長文の情報が入手可能な新聞社が配信する情報に対しても分析を行うとともに,長文の分類に際して低下すると思われるテキスト分類の精度向上を目指す.

参考文献

1) 齋藤勝久,近藤伸也,目黒公郎：子どもの防犯データベース設計に関する研究,生産研究, Vol. 61, No. 4 pp. 722-725, 2009.

2) 警察庁：平成２１年上半期の犯罪情勢警察統計, 2009.

表 4　警察統計と安全・安心メールの罪名件数比較

被害者	小学生		中学生	
媒体(件)	警察統計	メール	警察統計	メール
殺人	16	1	4	0
強盗	3	0	12	1
強姦	20	0	46	0
暴行	284	68	590	22
傷害	148	1	722	1
脅迫	7	0	29	0
恐喝	54	0	349	0
窃盗犯	10199	8	24885	30
詐欺	4	3	25	0
わいせつ	371	456	266	287
逮捕監禁	0	0	6	0
略取誘拐	23	0	9	0
その他	299	879	1272	275
計	**11428**	**1416**	**28215**	**616**