

被災経験マイニングに向けた Blog フィルタリング

Blog Filtering for Disaster Experience Mining

安部 智也†
Tomoya Abe

安藤 一秋‡
Kazuaki Ando

1. はじめに

近年、日本国内において大規模地震が多発し、多くの被害を引き起こしている。また、今後の大規模地震に伴う被害も懸念されており、防災対策の重要性が益々高まっている。各省庁や地方自治体では、ハザードマップの作成や緊急地震速報の整備等、様々な防災対策を実施・検討している。防災対策を実施・検討するにあたり、被災者の意見や感想・経験などの被災経験が情報の一部として活用されている。各省庁や地方自治体、企業などでは、アンケート調査[1]によって被災経験を収集しているが、アンケートの配布・集計は労力を要するため、回収率やコストに見合った情報が必ずしも得られないなどの問題がある。

これらの問題に対し、個人の経験や感想などが書かれている Blog を活用することで、不足する情報量を補うことができると考えられる。しかし、Blog 上の被災経験を検索エンジンで調べようとしても、Blog は構造化されていないため情報が分散しており、検索エンジンを用いても、効率的に被災経験を収集できない。

そこで本研究では、Blog から有用な被災経験をマイニングすることで、防災に役立つ有用な知識を構築することを目的とする。本稿では、まず研究の事前調査として、地方自治体や企業が実施している地震被災者に対するアンケート項目を整理し、Blog からの抽出可能性について分析する。そして、Blog から抽出できる項目の有用性について、専門家にヒアリングした結果を述べる。最後に、収集した Blog 集合から被災経験が記載されていない Blog をフィルタリングする手法と今後について検討する。

2. アンケート項目の調査

マイニングの対象を整理するため、地方自治体や企業が実施している地震被災者に対するアンケート 11 件に対し、共通の項目に分類して内容を調査する。なお、各アンケートの平均項目数は 9.2 件である。

調査の結果、アンケート項目は以下の 5 つに大別できた。

- ・地震発生時の行動 27 件
- ・防災対策 20 件
- ・回答者の属性 17 件
- ・地震に対する心理調査 10 件
- ・意見調査 6 件

アンケートには、このような被災者の意見や感想・経験などの被災経験を収集する項目が設定されている。上記の分類以外には、緊急地震速報の認識や評価など、特定の防災対策に関する項目も存在した。被災経験者から得られる

情報は、現在実施している防災対策の改善や新たな対策の検討などに対し、非常に有用であると考えられる。

3. Blog の調査と専門家に対するヒアリング

3.1. Blog とアンケート項目の関連性調査

2 で述べたアンケート項目が実際に Blog に記載されているかを予備調査する。調査対象として、岩手・宮城内陸地震の地震発生日から 3 日間の Blog を Yahoo! ブログ検索で 180 件収集し、分類する。なお、検索キーワードは「地震」を利用し、収集した Blog を人手で分類する。

分類結果を表 1 に示す。表 1 の項目の内、最も件数が多い「ニュースを聞いての感想」に対し、アンケート項目と一致する Blog 内容を調査したところ、被災者の意見や感想・経験が書かれていなかった。次に件数が多い「揺れを感じた」に対し、同様の調査を行った。

調査結果を表 2 に示す。どの項目も件数はあまり多くないが、Blog 記事からアンケート項目に関する情報が抽出できる可能性があることがわかった。

表 1 Blog 内容の分類結果

項目	件数
ニュースを聞いての感想	99 件
揺れを感じた	52 件
被災者への安否確認	20 件
過去の経験	5 件
その他	4 件

表 2 類似する内容調査

アンケート項目	アンケート項目と一致する Blog 内容	Blog 件数
意見調査	緊急地震速報の受信状況	6 件
地震発生時の行動	避難行動の方法	3 件
回答者の属性	住居の築年数	1 件
防災対策	住居の地震対策	1 件
地震発生時の行動	身内・友人への安否報告	1 件

3.2. 専門家へのヒアリング

マイニングの対象を整理するために、防災の専門家である長谷川修一教授（香川大学工学部）にヒアリングを行った。その結果、以下の 3 項目について有用性が確認できた。

1. 地震発生時の行動・心理状況
2. 地震後の行動・心理状況
3. 安否確認の方法

本研究では、地方自治体のアンケート項目と専門家のヒアリング結果で共通する地震発生時および地震後の行動と心理状況に注目する。行動と心理状況は、その場の状況に依存する内容であるため、本稿では、上記の 1 と 2 に加え、その場の状況について書かれた内容を被災経験と定義する。

以下に被災経験情報の図 1 に示す。

†香川大学大学院工学研究科 Graduate school of Engineering, Kagawa University

‡香川大学工学部 Faculty of Engineering, Kagawa University

自分は海外家具の店、「IKEA 新三郷店」に嫁と 2 人で行ってりました。家具が大量に積まれているお店で、今回のような規模の地震が突然来た場合、かなり恐怖です。揺れ始め数秒後、店内ではガシャンッ、ガシャンッと何かの割れる音と お客の悲鳴や怒号が響き渡りました。店内放送が響き渡り、店員の誘導で全員店の外へ。

<http://amaebi.net/archives/1705196.html>(2012 年 3 月 11 日)

図 1 被災経験情報の例

3.3. 地震について書かれた Blog の予備調査

地震について書かれた Blog の存在を簡易的に確認するため、東日本大震災発生日を含む、2011 年 1 月 1 日から 6 月 31 日の期間に記述された Blog 件数の推移を調査する。地震について書かれた Blog には、本文中に「地震」というキーワードが存在する仮定し、Yahoo! ブログ検索を用いて記事数の調査を行う。

検索結果を図 2 に示す。図 2 より、地震発生後は、急激に「地震」を含む Blog の件数が増加していることが分かる。以上より、平常時に比べて Blog 記事数の推移に変動が大きい期間に書かれた Blog には、被災経験が書かれている可能性があると考えられる。

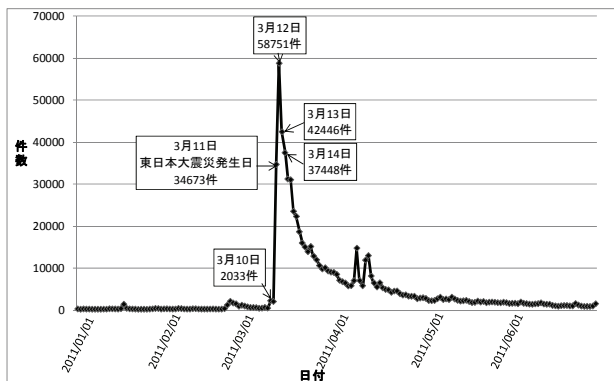


図 2 地震について書かれた Blog 記事数の推移

4. 被災経験を含まない Blog のフィルタリング

キーワード検索では、被災経験の有無を考慮して Blog 記事を収集できない。そこで、「地震」というキーワードを含む Blog 記事を網羅的に収集した後、その集合から被災経験を含まない Blog をフィルタリングする手法について検討する。フィルタリング手法については様々な方法が提案されているが、本研究では代表的なフィルタリングであるベイジアンフィルタと SVM (Support Vector Machine) を比較し、性能の高い方を利用する。

5. フィルタリング手法の性能評価

ベイジアンフィルタと SVM の性能を評価し、フィルタリングに採用する手法を検討する。評価データは、新潟県中越沖地震 656 件と静岡沖地震 722 件を利用する。各データを人手で分類した結果を表 3 に示す。

SVM の実装には、SVM-Light を利用し、仕様としてはいくつか検討した結果、線形カーネルを採用した。SVM の素性には、単純に本文に含まれる名詞と動詞の形態素を利用し、素性の重みは TF-IDF を利用する。また、形態素

の複合名詞化を行い、有効性についても調査する。また、正例・負例データの割合による影響を受けにくい手法を選択したいので、学習データの割合を変更した 3 つのデータセットを用いる。評価方法は、各データセットに対して、10 点交差検定を行い、各手法に対する精度・再現率・F 値の平均値で比較する。

評価結果を表 4 に示す。表 4 に示すように、各手法の F 値にはあまり差が見られなかったが、ベイジアンフィルタが若干高い値を示した。本研究の最終目的は経験情報のマイニングであるため、経験情報が記載された Blog を漏れなく収集できる手法が望ましい。SVM については、素性として Bag-of-Words しか利用していないため、今後は他の素性について検討し、総合的に目的が達成できるフィルタリング手法を採用する予定である。

表 3 学習データ

	新潟県中越沖地震	静岡沖地震
正例データ	154 件	401 件
負例データ	502 件	321 件

表 4 フィルタリングの結果

機械学習	複合名詞	精度	再現率	F 値
SVM	なし	78.033	73.213	74.608
	あり	76.323	70.717	72.330
ベイジアン フィルタ	なし	66.370	87.596	75.498
	あり	69.514	85.488	75.948

6. 今後の方針

本研究では、その場の状況と行動および心理状況を経験情報とみなしてマイニングする。これらの経験情報は、1 文単位で取り出すより、Blog 記事中からパッセージとして取り出すことが望ましいと考えられる。今後は、1 文単位で取り出す方法とパッセージとして取り出す手法などを検討する予定である。そして、経験文またはパッセージからマイニングした結果の表現形式についても検討する。

7. おわりに

本研究では、Blog から有用な被災経験情報をマイニングすることが目的である。本稿では、まず地方自治体や企業が実施している地震被災者に対するアンケート項目と Blog の内容の関係を調査し、Blog からアンケート項目に関する回答が抽出できる可能性を示した。また、専門家にヒアリングすることで、マイニング対象をその場の状況と行動および心理状況に決定した。さらに、地震発生日と「地震」を含む Blog が急激に増加することを示した。そして、Blog から有用な被災経験をマイニングするための前段階として、被災経験を含まない Blog をフィルタリングする手法について検討した。

今後は、フィルタリングの性能を向上させると共に、経験情報を取り出す手法などについて検討する。

参考文献

[1] 内閣府, 平成 16 年新潟県中越地震に関する住民アンケート調査, <http://www.bousai.go.jp/km/skk/data/data02-03.pdf>