1. はじめに

Twitter など SNS の利用者は爆発的に増えており，それに伴い不適切な投稿内容や写真が第三者により発見・拡散され，炎上する事件（通称“バカッター”等）が増えている。炎上の発端となった投稿内容がマスコミで報道されれば，企業イメージ悪化に伴う損害賠償請求など，取返しのつかない事態に発展してしまう。

図１に典型的な炎上プロセスを示す。不適切な投稿等に対して,投稿から 30 分ほどで第三者が批判コメントを投稿し，2 時間でネット上に投稿が拡散する。2 日以内に個人情報が特定され，それ以降，メディアで報道され，大事件化することも多い。

そこで，SNS での炎上被害などを減らし，安全に利用するためのツール「SNS 炎上報知器」を提案する。

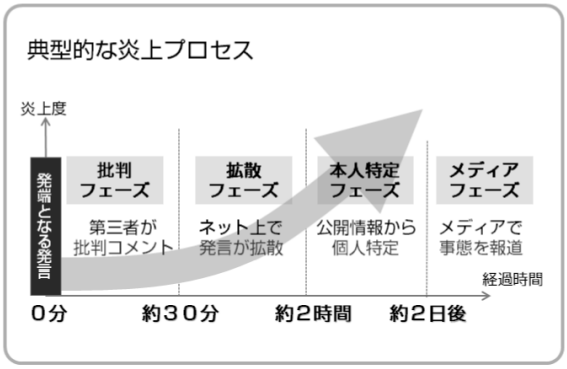

図１　典型的な炎上プロセス

2. システム概要（図２）

「SNS 炎上検知器」は，Android スマートフォン上で使用し，投稿内容やリツイート・リプライ数を監視することでユーザーに”気づき”の機会を与え，炎上を事前に予防又は対策できるようにする。図１の炎上の各フェーズに対応し，次の機能を開発した。

2.1 批判フェーズ（NG ワード・個人情報の検出）

投稿直後 30 分以内の批判フェーズでは,投稿内容に犯罪予告や誹謗中傷などの NG ワードや個人情報が含まれていないか，を検出する。投稿内容に対してリアルタイムに検出を行い，ユーザーに削除・謝罪を促すことで，拡散フェーズへの移行を防ぐ。

2.2 拡散フェーズ（炎上感知）

拡散フェーズでは， Twitter や Facebook のリツイート，リプライ，いいね，シェアの一定時間内の増加数をチェックし，異常に増加している場合には炎上の可能性を警告する。

2.3 本人特定フェーズ（事前予防）

本人特定フェーズに移行してしまうと，個人の特定は避けられない。そこで事前予防として定期的にプロフィール等の公開設定をチェックし設定の不備に対して，ユーザーに”気づき”を与える。また一定期間内の Twitter と Facebook の投稿内容から個人情報を特定できる可能性を統計処理・特徴抽出により数値化する。

図２　システム概要

3. 将来のビジネスプラン

将来のビジネスプランとしてはフリーミアム方式を検討している。「SNS 炎上検知器」は一定のユーザー数と評価を獲得するため，一般ユーザーには無償提供する。法人向けには社員や所属タレントの投稿をサーバで一括管理する機能を提供することで収益化を目指す。

4. まとめ

「SNS 炎上検知器」は Android スマートフォン上で，投稿内容や公開設定等をリアルタイムに監視し，ユーザーに“気づき”の機会を与えることで炎上を予防又は対策する。将来は法人向けの機能を追加し，収益化を図りたい。