실시간 SNS 스트림 처리를 통한 산림재해 관련 메시지 검출 시스템 구현

Implementation of Forest Disaster Message Detection System by Processing Real-time SNS Stream

문 지 혜, 임 지 수, 송 석 일*
한국교통대학교

Jihye Moon, Jisu Lim, Seokil Song*
Korea National University of Transportation

요약

이 논문에서는 SNS 스트림을 분석하여 산림재해 (산불, 산사태 등)에 관련한 메시지를 검출하는 시스템을 구현한다. 구현하는 시스템은 SNS 스트림에서 사전에 부여된 산림 재해 과련 키워드들을 포함하는 메시지를 실시간으로 검색한다. 검색 결과에 대해 산림재해와의 관련성을 관리자가 분류를 하고 분류한 결과는 기계학습 기법을 통해 학습되어 보다 정확한 산림재해 메시지 검출을 가능하게 한다. 제안하는 방법은 Spark Streaming 과 MLIB를 기반으로 구현한다.

I. 서론

산림재해를 조기에 발견하고 이를 다양한 매체를 통해서 알리는 것은 산림재해를 통한 피해를 줄이는데 있어 매우 중요하다. 기존에 [1-3] 과 같은 산림재해를 예측하고 발견하기 위한 다양한 ICT융합기술이 제안된바 있다. 대부분 영상정보나 센서, 기상 정보를 이용한 산림재해의 예측 또는 검출이 주 내용이다.

이 논문에서는 SNS 스트림 분석을 통해서 산림재해가 (산불, 산사태 등) 발생했거나 산림재해로 추측되는 메시지를 검출하는 시스템을 구현한다. 구현하는 시스템은 SNS 스트림에서 직접적으로 산림재해를 언급하는 메시지를 검출하거나 또는 산림재해를 간접적으로 유추할 수 있는 메시지를 검출한다.

SNS 스트림에 대한 산림재해에 관련된 단순 키워드 검색으로는 오경보(false alarm)을 막을 수 없다. 즉, '산불', '산사태' 등의 산림재해 키워드를 SNS 스트림에 대해 실시간으로 검색을 수행하면 그림 1에서와 같이 오경보가 발생하게 된다. 즉, '산불' 키워드를 포함하고 있지만, 실제로는 산불 조심 캠페인인 경우가 있을 수 있다.

이 논문에서는 이와 같은 오경보를 방지하기 위해서기계학습기법을 적용한다. 초기에는 단순 키워드 검색을 통해서 추출된 산림재해 후보 메시지들에 대해서 관리자에 의해 분류가 진행된다. 분류된 메시지들은 SVM (Support Vector Machine) 과 같은 분류기법의 학습 데이터로 사용되게 되며 학습을 통해 구축된 분류 모델을 통해서 후보 메시지들을 분류하여 산림재해 메시지들을 검출하여 오경보를 줄이게 된다.

▶▶ 그림 1. 오경보 사례

2장에서는 이 논문에서 구현한 산림재해 SNS 메시지 검출 시스템에 대해서 기술하고 3장에서는 결론을 맺는다.

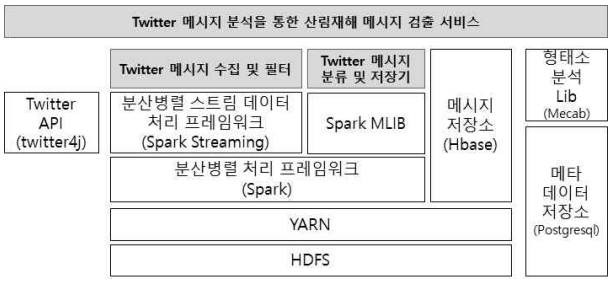
II. 산림재해 SNS 메시지 검출 시스템 구현

이 논문에서 설계하고 구현하는 SNS 스트림에서의 산림재해 메시지 검출 시스템의 구조는 그림 2와 같다. 그림에서와 같이 개발하는 메시지 검출 시스템은 Hadoop 및 Spark[4]을 기반으로 하고 있다. Spark의 Spark

* 교신저자 : sisong@ut.ac.kr
본 연구는 2015년 산림청 임업기술연구개발사업의 지원을 받아 수행한 연구 과제(과제번호 :S111515 L010130)입니다

Streaming을 이용하여 관리자에 의해서 주어진 키워드들을 중심으로 SNS 메시지 스트림에서 산림재해 후보 메시지를 실시간으로 추출한다.

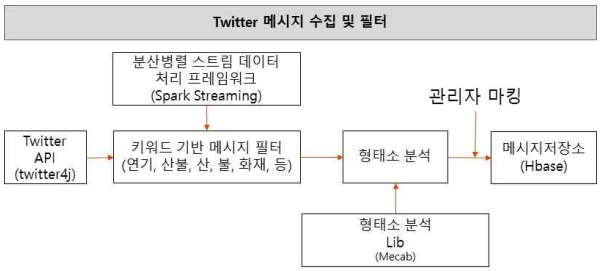
추출된 산림재해 관련 메시지들은 후보 메시지이며 후보 메시지들은 초기에는 관리자에 의해 수동으로 분류된다. 분류된 산림재해 및 비 산림재해 메시지들은 형대소 분석기에 의해 단어 벡터로 변환되어 메시지 저장소에 저장된다. 일정 개수 이상의 후보 메시지가 저장되면 기계학습 알고리즘 (이 논문에서는 SVM)을 이용하여 학습을 진행한다. 학습을 통해 분류 모델이 도출되면 이후에 추출되는 후보 메시지에 대해서는 SVM 분류기를 통해서 자동으로 분류를 수행한다.



▶▶ 그림 2. 구현한 산림재해 메시지 검출 시스템 구조

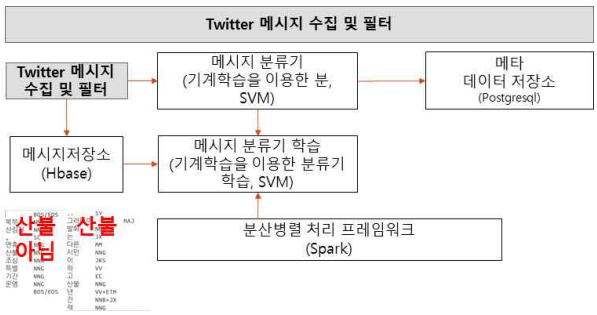
이 논문에서는 트위터 SNS를 이용하여 시스템을 구현하였다. 그림 3은 트위터 메시지 스트림에 대해 실시간으로 키워드 검색을 수행하여 산림재해 후보 메시지를 추출하고 관리자에 의해 분류를 수행한 후 형태소 분석을 통해 단어벡터로 전화하여 메시지저장소에 저장하는 흐름을 보여주고 있다.

그림에서처럼 트위터 메시지 스트림에 대한 실시간 처리는 스트림 처리기인 Spark Streaming을 이용하고 있다. 또한, 형태소 분석은 Mecab이라는 형대소 분석 라이브러리를 사용한다.



▶▶ 그림 3. SNS 스트림에서 산림재해 후보 메시지 추출 과정

그림 4는 앞 단계에서 추출한 후보 메시지를 메시지저장소에 기계학습을 통해서 학습하고 이를 통해 후보메시지를 자동으로 분류하는 과정을 보여주는 그림이다. 이 논문에서는 기계학습 알고리즘으로 SVM을 이용하며 Spark에서 제공하는 MLib를 기반으로 구현한다.



▶▶ 그림 4. 기계학습을 기반으로 하는 후보메시지 분류 과정

Ⅲ. 결론

이 논문에서는 SNS 중 하나인 트위터 스트림을 실시간으로 분석하여 산림재해 (산불, 산사태 등)에 대해 직접적으로 언급하거나 산림재해를 유추할 수 있는 메시지를 검출하는 시스템을 설계하고 구현하였다. 대용량의 트위터 스트림 메시지를 처리하기 위해서 분산 병렬처리 플랫폼인 Hadoop 및 Spark을 기반으로 설계하고 구현하였다. 이 논문에서는 오경보를 줄이기 위해서 기계학습 기법을 도입하였다.

향후 연구에서는 구현한 시스템에 대한 성능평가를 수행할 계획이다. 성능평가 지표로는 분류의 정확도를 선정할 계획이다. 이를 통해서 이 논문에서는 구현한 시스템이 산림재해 메시지 검출시 오경보를 얼마나 줄일수 있는지 확인한다.

■ 참 고 문 헌 ■

[1] 이승희, 신범주, 송복득, 안선정, 김진동, 이학준, “영상 접합을 이용한 산불 감시 시스템,” 한국콘텐츠학회논문지 제13권 제6호, pp.40-47, 2013.

[2] 김형준, 신규영, 우병훈, 구남경, 장경식, 이강환, “상황인지 센서를 활용한 지능형 산불 이동 예측 및 탐지 알고리즘에 관한 연구,” 한국정보통신학회논문지 제19권 제6호, pp.1506-1514, 2015.

[3] 안명수, 강대성, “기상 및 비전 센서 기반의 산불 방재 시스템에 관한 연구,” 한국정보기술학회논문지 제12권 제6호, pp.175-180, 2014.

[4] Zaharia, Matei, et al. “Spark: Cluster Computing with Working Sets.” HotCloud 10, 2010