

## 텍스트 기반 119 신고전화 상황 분류

### A text-based emergency situation classification method

곽 세 민\*·임 윤 섭\*\*·최 중 석\*\*\*

Semin Kwak, Yoonseob Lim, JongSuk Choi

#### 요 약

본 논문에서는 기계학습 방법에 기반을 둔 119 긴급 신고 전화 전사 데이터에 대한 구급, 구조, 화재 상황 분류 알고리즘을 개발하였다. 신고전화에서 빈번하게 발생하는 비정형 발화 패턴을 효율적으로 정규화하고 자연어 문장 처리 기법에서 일반적으로 사용하는 방법을 적용하여 신고전화 텍스트 데이터를 기계학습에서 사용할 수 있는 특징 벡터로 재구성하였다. 2743개의 신고전화에 대해 선형 서포트 벡터 머신을 이용하여 상황 분류를 수행한 결과, 92%의 정확도를 얻을 수 있었다.

**keywords** : 서포트 벡터 머신 (support vector machine), 선형 최적화, 기계 학습

## 1. 서 론

재난 상황의 빠르고 정확한 판단과 대응은 119 수보자의 대응 매뉴얼의 첫 번째에 등장 할 만큼 효과적인 신고전화 응대를 위한 가장 중요한 요소 중의 하나이다. 그럼에도 불구하고 현재 신고전화 대응 시스템에서는 온전히 119 수보자의 상황 판단 능력에 의존하고 있다. 본 논문에서는 신고상황 판단 과정을 보다 효율적으로 수행하기 위해서 기계학습 방법을 이용하여 119 신고전화에서 일반적으로 발생하는 구급, 구조, 화재 상황에 대한 분류가 가능함을 보여준다. 기계학습 과정에 사용된 모든 데이터는 실제 119 전화를 기반으로 전사된 문장 데이터이며, 학습 데이터와 테스트 데이터를 완전히 분리하여 학습 및 검증 과정에 적용하였다.

## 2. 본론

화재, 구급, 구조 상황의 비정형 문자를 분류하기 위하여 미리 상황이 분류되어 있는 총 3351개의 실제 119 신고 전화 문장 데이터 (화재 298, 구급 2224, 구조 829개)에 대하여 기계학습을 수행하였다. 기계학습의 수행에서는 608개의 문장 데이터 (화재 87, 구급 381, 구조 140개)에 대하여 학습에 사용하였고, 나머지 2743개의 문장 데이터 (화재 211, 구급 1843, 구조 689개)는 학습된 모델에 대한 성능 검증에 사용하였다.

### 2.1. 문장 데이터 전처리 (Text preprocessing)

\* 한국과학기술연구원 로봇연구단 연구원 semin.kwak@kist.re.kr

\*\* 한국과학기술연구원 로봇연구단 선임연구원 yslim@kist.re.kr

\*\*\* 한국과학기술연구원 로봇연구단 책임연구원 cjs@kist.re.kr

119 신고 전화의 특성상 대부분의 신고자는 신고지의 주소를 언급한다. 주소 정보는 신고 전화마다 고유한 값을 갖게 되므로 기계학습 과정에서 과적합 (overfitting) 문제를 야기할 수 있다. 과적합 문제를 회피하기 위해 다양한 방법에 기계학습에서 적용되고 있으며, 본 연구에서는 신고자의 주소정보를 공통된 문자 (주소태그)로 치환함으로써 과적합 문제를 해결하였다. 또한 형태소 분석기를 사용하여 신고전화 문장 데이터에서 핵심 형태소인 주어, 동사만을 사용하였다. 표 1. (가)는 문장 데이터 전처리가 실제로 어떻게 수행되었는지를 보여준다. 예를 들어 본 논문에서 제시한 전처리 과정에서 주소정보 ‘의정부’가 ‘주소태그’로 변경됨을 알 수 있다.

표 1 기계학습의 문장 데이터 전처리 및 특징 추출 예

| 원본 텍스트                       | 텍스트 전처리                        | 형태소 추출         | 특징 추출 벡터                                 |
|------------------------------|--------------------------------|----------------|--|
| 지금 여기<br>의정부인데요<br>산에서 불이 나요 | 지금 여기<br>주소태그 인데요<br>산에서 불이 나요 | 주소태그 산 불<br>나다 | [1 0 0 ... 0 1 0<br>... 1 0 ... 1 ... 0] |
| (가) 문자 전처리                   |                                |                | (나) 특징 추출                                |

## 2.2. 특징 벡터 추출 (Feature vector extraction)

자연어 처리 기법에서 일반적으로 사용되는 방법을 이용하여 전처리된 신고전화 데이터에 대해 특징 벡터로 재구성하였다. 본 논문에서는 학습에 사용되는 모든 문장 데이터를 구성하는 단어들의 출현 빈도수에 기반하여 상위 3000개의 단어를 추출하고 각 단어별로 그에 따른 순위를 배정하였다. 각 신고전화 문장데이터는 길이 3000의 벡터로 구성되며, 해당 신고전화에서 발생한 단어의 순위에 1이라는 값을 부여하는 방식으로 특징벡터를 구성하였다. 예를 들어 [지금 불이 나고 있습니다.] 의 콜은 단어-숫자 할당 표에서 ‘지금’, ‘불이’, ‘나고’, ‘있습니다’ 의 숫자의 순위를 찾는다. 만약 ‘지금’-1, ‘불이’-3, ‘나고’-4, ‘있습니다’-6 으로 숫자가 할당 되어 있었다면 이 콜의 특징 벡터는 [1 0 1 1 0 1 ... 0]이 된다.

## 2.3. 학습

본 연구에서는 모든 콜의 특징벡터-상황 쌍 (화재 87콜, 구급 381콜, 구조 140콜)을 선형 서포트 벡터 머신 기법을 이용하여 기계학습을 수행하였다. 서포트 벡터 머신 기법은 선형 모델로써 모든 식이 행렬과 벡터로 표현 가능하여 빠르고 안정적인 학습이 가능하다.

## 3. 결론

긴급 상황 분류를 위하여 학습에 사용되지 않은 2743개의 신고전화 데이터 (화재 211건, 구급 1843건, 구조 689건)를 학습된 모델을 이용하여 분류하고 그 성능을 검증하였다 (표. 2). 분류 성능은 기계학습에서 사용되는 정밀도, 재현율 및 F1-Score로 나타내었다. 정밀도는 모든 분류된 신고전화 상황 중에서 실제 맞게 분류된 상황의 비율이며, 재현율은 모든 신고전화 중 실제 정상적으로 상황이 분류된 것의 비율이다. F1-score는 다음과 같다.

$$F1 - score = 2 \frac{\text{정밀도} \times \text{재현율}}{\text{정밀도} + \text{재현율}}$$

표 2에서 나타나듯이 본 논문에서 제안하는 상황분류기의 성능은 평균적으로 0.92의 F1-score를 나타내었다.

표 2 긴급 상황 분류기 성능

|    | 정밀도  | 재현율  | F1-score    | 콜 개수 |
|----|------|------|-------------|------|
| 화재 | 0.89 | 0.89 | 0.89        | 211  |
| 구급 | 0.93 | 0.97 | 0.95        | 1843 |
| 구조 | 0.92 | 0.83 | 0.87        | 689  |
| 평균 | 0.93 | 0.93 | <b>0.92</b> | 2743 |

### 감사의 글

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음. [R0126-16-1119, 음성·음향 분석 기반 상황 판단 솔루션 기술 개발]

### 참고문헌

- Christopher D Manning.** (1999) Foundations of Statistical Natural Language Processing, MIT Press.
- JAK Suykens.** (1999) Least squares support vector machine classifiers, Neural processing letters, 9,3 (1999): pp.293~300.