

스팸 필터링을 위한 지식 그래프 기반의 신조어 감지 매커니즘[☆]

Knowledge Graph-based Korean New Words Detection Mechanism for Spam Filtering

김 지 혜¹ 정 옥 란^{*}
Ji-hye Kim Ok-ran Jeong

요 약

오늘날 스마트폰에서 스팸 문자를 차단하기 위해 문자 내용과 스팸 키워드의 단순 문자열 비교 또는 스팸 전화번호를 차단하는 방식을 사용하고 있다. 이에 따라 스팸 문자가 자동으로 차단되는 것을 방지하기 위해 점차 변화된 방식으로 스팸 문자를 전송한다. 특히 스팸 키워드에 포함되는 단어의 경우 단순 문자열 비교로 검색되지 않도록 특수문자, 한자, 띄어쓰기 등을 이용하여 비정상적인 단어로 스팸 문자를 발송한다. 기존 스팸 필터링 방식의 경우 이러한 스팸 문자를 차단할 수 없다는 한계가 있다. 따라서 변화하는 스팸 문자에 대응할 수 있는 새로운 기술이 필요한 시점이다. 본 논문에서는 스팸 문자에서 자주 사용되는 신조어를 검출하여 변화하는 스팸 문자에 대응할 수 있는 지식 그래프 기반의 신조어 감지 매커니즘을 제안한다. 또한 기본 Naive Bayes에 감지한 신조어를 적용하여 제안한 방법의 성능 실험 결과를 보여준다.

☞ 주제어 : 스팸 필터링, 스팸 탐지, 신조어 감지, 지식그래프, 컨셉넷

ABSTRACT

Today, to block spam texts on smartphone, a simple string comparison between text messages and spam keywords or a blocking spam phone numbers is used. As results, spam text is sent in a gradually changed way to prevent it from being automatically blocked. In particular, for words included in spam keywords, spam texts are sent to abnormal words using special characters, Chinese characters, and whitespace to prevent them from being detected by simple string match. There is a limit that traditional spam filtering methods can't block these spam texts well. Therefore, new technologies are needed to respond to changing spam text messages. In this paper, we propose a knowledge graph-based new words detection mechanism that can detect new words frequently used in spam texts and respond to changing spam texts. Also, we show experimental results of the performance when detected Korean new words are applied to the Naive Bayes algorithm.

☞ keyword : Spam Filtering, Spam Detection, New Words Detection, Knowledge Graph, ConceptNet

1. 서 론

오늘날 많은 회사들이 마케팅을 위해 휴대폰으로 광고성 문자를 모방하여 정상 문자와 유사한 스팸 문자를 발송한다. 이러한 스팸 문자로 인해 회사 사칭에 따른 특정

회사의 이미지 손실 또는 스미싱으로 인한 고객의 금전적 피해 등 여러 사회적 문제가 발생한다.[1,2] 또한 스팸 문자로 차단되지 않기 위해 문자 내용에 특수문자, 한자, 기호 등 의미 없는 문자를 추가하여 비정상적인 형태로 스팸 문자를 발송한다.[3] 특히 알파벳 또는 특수문자를 이용하여 한글, 숫자 또는 URL 등을 표시할 경우 기존의 스팸 키워드 방식으로는 스팸 문자를 분류하기 어렵다.[4,5]

현재의 스팸 문자를 분류하기 위해서는 스팸 필터링 시스템이 사람과 비슷한 수준으로 문자의 내용을 이해해야 한다. 그러나 휴대폰 문자는 글자 수 제한으로 인해 이메일에 비해 짧은 문장으로 이루어져 있다.[6] 이처럼 짧은 텍스트에서 문장의 의미를 파악하기 위해서는 스팸 문자에서 사용되는 단어 정보를 파악하는 것이 중요하다.

본 논문에서는 변화하는 스팸 문자에 대응하기 위해

¹ Dept. of Software, Gachon University, Gyeonggi-do, 13120, Republic of Korea

* Corresponding author: orjeong@gachon.ac.kr

[Received 9 January 2020, Reviewed 30 January 2020, Accepted 14 February 2020]

☆ This research was supported by Basic Science Research Program through the NRF(National Research Foundation of Korea), and the MSIT(Ministry of Science and ICT), Korea, under the National Program for Excellence in SW supervised by the IITP(Institute for Information & communications Technology Promotion) (Nos.NRF 2019R1A2C1008412, 2015-0-00932).

스팸 문자에서 볼 수 있는 신조어를 지식 그래프를 이용하여 감지하는 새로운 신조어 감지 매커니즘을 제안한다. 이렇게 감지한 신조어로 지식 그래프를 구축하여 확장 가능한 스팸 키워드 시스템을 구현할 수 있다.

2. 관련 연구

2.1 지식 그래프 및 ConceptNet

지식 그래프는 단어와 단어 사이의 관계 정보를 표현하기 위해 서로 관련된 단어 별로 연결한 다관계 그래프이다.[7] 잘 구축된 지식 그래프에서는 그래프 노드 간의 새로운 관계 정보를 찾을 수 있기 때문에 텍스트 기반의 관계 추출에 중요한 보조 수단으로 사용된다.[8]

ConceptNet은 오픈소스로 공개된 지식 그래프로서 다양한 언어를 지원하는데, 10개의 핵심 언어와 한국어를 포함한 68개 이상의 공통 언어로 구성되어 있다. 핵심 언어로는 영어, 프랑스어, 일본어, 중국어 등이 있으며, 핵심 언어의 경우 지식 그래프 내에 최소 24만 개 이상의 어휘를 갖고 있다. 특히 영어의 경우 180만 개 이상의 어휘로 지식 그래프가 구성되어 있어 규모가 매우 크다.[9] 영어에 대한 ConceptNet 5의 성능을 살펴보면, 임의의 샘플 데이터에 대해 참가자들이 정답 여부를 분류한 실험에서 WordNet 3.0보다 더 좋은 평가를 받았다.[10]

그러나 이에 비해 공통 언어의 경우 구축되어 있는 그래프 내의 어휘 수가 매우 부족하여 추가적인 학습 없이 사용하기 어렵다. 특히 한국어의 경우 약 47,000개 어휘만을 갖고 있어 영어와 비교해 볼 때 매우 부족한 수준이다.

2.2 스팸 필터링

스팸 문자를 탐지하는 방법으로 크게 두 가지 방식이 사용된다. 협업 기반 방식과 콘텐츠 기반 방식이다. 협업 기반 방식의 경우 사용자들의 피드백과 공유된 사용자 경험에 기초하여 스팸 여부를 판단한다. 이와 달리, 콘텐츠 기반 방식은 텍스트 내용을 분석하여 스팸을 분류하는 방식이다. 사용자 경험에 대한 데이터를 얻는 것은 매우 어렵기 때문에 대부분의 스팸 필터링 관련 연구는 콘텐츠 기반 방식을 사용한다.[11]

초기의 스팸 문자 탐지에는 규칙 기반, Naive Bayes 방식 및 SVM (Support Vector Machines) 방식을 사용하였다.[12-14] 이후 k-nearest neighbor algorithm (k-NN) 및 Naive Bayes (NB) 등의 방식이 사용되었다.[15,16] Jie et al

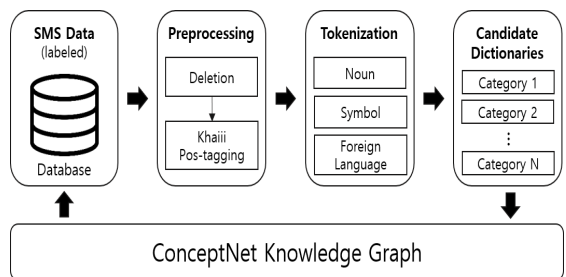
에서는 거짓 양성 (false positive)을 높이기 위해 단어에 가중치를 더하는 방식이 사용되기도 하였다.[17]

최근에는 deep learning을 적용한 CNN 기반 스팸 필터링 방식과[18-20] SentiWordNet을 이용한 sentiment analysis 기반의 스팸 필터링[21], 유사도 기반의 말뭉치 및 위키피디아 링크 기반의 스팸 필터링[22] 등 다양한 스팸 필터링 방식이 제안되고 있다.

3. 스팸 신조어 감지 매커니즘

본 논문에서 제안하는 지식 그래프 기반 스팸 신조어 감지 매커니즘은 그림 1과 같이 크게 스팸 데이터의 전처리, 토큰화, 후보 사전 구축, 지식 그래프 검색 순으로 4단계로 이루어진다. 전처리 단계에서는 스팸 여부와 관계없는 문자열을 삭제한다. 이렇게 정제된 스팸 문자에서 토큰화 단계를 거쳐 명사, 기호, 외국어 세 가지 타입의 문자만 토큰으로 추출한다. 추출된 토큰으로 카테고리 별 신조어 후보 사전을 구축한 후 상위 N개의 토큰에 대해 지식 그래프 검색 결과를 바탕으로 신조어를 감지한다.

신조어 감지를 위해 데이터베이스로부터 스팸의 카테고리별 스팸 여부가 라벨로 구분된 스팸 데이터를 가져온다. 가져온 스팸 데이터에 대해 전처리 없이 신조어를 감지하는 방법도 가능하지만, 이 경우 스팸 문자에서 단어를 추출하는데 어려움이 많다. 이를 해결하기 위해 본 논문에서는 스팸 문자의 특성을 파악하여 전처리 및 토큰화를 진행한다.



(그림 1) 지식 그래프 기반 신조어 감지 매커니즘
(Figure 1) Knowledge Graph-based New Words Detection Mechanism

3.1 전처리

스팸 문자의 전처리는 삭제와 품사 태깅 두 단계로 이루어진다. 각 단계를 거쳐 기존 스팸 문자로부터 정제된

스팸 문자를 추출한다.

3.1.1 삭제 (Deletion)

먼저 스팸 문자의 전처리를 위해 “광고”와 같이 스팸 문자와 비스팸 문자 모두에 공통적으로 사용되는 문자열을 삭제한다. 또한 데이터를 수집하는 과정에서 포함되는 문자열을 삭제하는 과정을 거친다. 이 때, 따옴표, 줄바꿈($\backslash n$) 등의 불필요한 문자 역시 삭제한다.

3.1.2 품사 태깅 (Pos-tagging)

스팸 문자에서 단어를 추출하기 위해 Kakao에서 개발한 세 번째 형태소 분석기인 khaiii를 이용하여 품사를 태깅한다. 각 스팸 문자의 문자열은 khaiii 사전에 의해 총 46개의 품사 중 하나로 태깅되며 khaiii 사전에 원하는 단어를 추가할 수 있다.

3.2 토큰화

스팸 문자를 단어 토큰 단위로 분리하기 위해 형태소 분석기의 명사 추출 라이브러리를 이용할 수 있다. 그러나 라이브러리를 이용할 경우 손쉽게 명사를 추출할 수 있다는 장점이 있으나 추출되는 단어의 수가 극히 적다는 단점이 있다. 더불어 토큰화에 사용하는 라이브러리의 단어 사전에 존재하지 않는 단어의 경우 토큰으로 추출할 수 없다는 문제가 있다. 따라서 본 논문에서는 품사를 이용하여 단어를 추출하는 과정을 거친다.

전처리가 완료된 각각의 스팸 문자를 토큰화할 때, khaiii의 46개 품사 중 명사(‘NNG’, ‘NNP’, ‘NNB’, ‘NP’, ‘NR’, ‘SWK’), 기호(‘SN’, ‘SW’), 외국어(‘SL’, ‘SH’) 품사에 해당하는 경우에만 단어로 추출하였다. 명사 뒤에 사용되는 조사와 동사, 부사, 어미, 접미사 등의 품사에 해당하는 문자열은 단어에 포함되지 않으므로 토큰 추출에 포함하지 않는다. 기호 중에서도 따옴표, 괄호표, 줄표를 의미하는 ‘SS’ 품사와 마침표, 물음표, 느낌표를 의미하는 ‘SF’ 품사, 쉼표, 가운데점, 콜론, 빗금을 의미하는 ‘SP’ 품사의 경우에도 토큰 추출에 포함하지 않는다.

3.3 후보 사전 구축

후보 사전은 스팸 카테고리 별 토큰과 각 토큰이 스팸 문자에서 사용된 횟수로 구축된다. 토큰화를 통해 추출된 토큰은 스팸 문자 내에서 사용된 횟수만큼 중복되어 나

타난다. 여기서 중복 토큰 수는 곧 단어의 빈도수를 의미한다. 따라서 토큰의 중복 처리 및 각 토큰의 빈도수 계산을 위해 신조어 후보 사전을 구축한다.

신조어 후보 사전은 스팸 카테고리마다 별도로 구축한다. 구축한 신조어 후보 사전은 빈도수가 높은 순으로 정렬하여 이후 지식 그래프 검색에서 사용한다.

3.4 지식 그래프 검색

내림 차순으로 정렬된 신조어 후보 사전에서 특정 임계값 이상 사용된 토큰에 대해 지식 그래프인 ConceptNet에 검색한다. 지식 그래프 검색 시, 검색 토큰의 연관 단어가 ConceptNet에 존재하지 않거나, 검색 결과가 기준치 이하로 적을 경우 해당 토큰을 신조어로 감지한다.

4. 실험

본 실험에는 ConceptNet의 Web API와 형태소 분석을 위해 Kakao의 Khaiii를 사용하였으며, 실험은 Google colab의 TPU 환경에서 진행하였다.

4.1 데이터셋

(표 1) 카테고리 별 스팸 건수

(Table 1) Number of Spam in each category

No.	Category	Number
1	도박	6,182
2	금융	2,877
3	선거	2,596
4	대리운전	1,680
5	비스팸	1,669
전체		15,004

실험에는 KISA에서 제공받은 15,004건 (총 224,121 단어)의 스팸 문자 데이터를 사용하였다. 스팸 문자 데이터는 ‘도박’, ‘선거’, ‘대리운전’, ‘금융’, ‘비스팸’으로 총 5개의 카테고리로 분류되어 있다. 표 1은 카테고리에 따른 스팸 문자 데이터의 개수를 나타낸다. 전체 스팸 카테고리 중 ‘도박’으로 분류된 스팸 문자가 6천 건 이상으로 가장 많으며 ‘금융’, ‘선거’ 순으로 전체 데이터에서 큰 비중을 차지한다.

4.2 신조어 감지

표 2는 토큰화에서 단어 추출을 위해 사용된 Khaiii 품사의 목록이다. 토큰이 명사, 기호, 외국어 품사에 해당할 경우 토큰들을 연결하여 하나의 단어로 처리하며 이외의 품사에 해당하는 토큰이 나타날 경우 이를 기준으로 후보 단어를 저장한다.

예를 들어 앞뒤 토큰의 품사가 명사+조사일 경우 명사만 후보 단어로 추출하며 외국어+'SS'(괄호표)일 경우 외국어만 후보 단어로 추출한다. 'SS'(괄호표)+명사+기호+'SP'(마침표)일 경우 명사+기호를 하나의 단어로 추출한다. 표 2에 없는 품사의 경우 단어 추출에 포함하지 않는다.

더불어 한 글자로 이루어진 토큰의 경우 의미를 파악하기 어렵기 때문에 명사, 기호, 외국어에 해당해도 토큰에서 제외하였다. 이러한 토큰화 과정을 통해 총 66,529개의 단어를 추출하였다.

(표 2) 토큰화에 사용된 Khaiii 품사 태그

(Table 2) Khaiii Pos tags used in Tokenization

Part of Speech	Khaiii tags	Explanation
명사	NNG	일반 명사
	NNP	고유 명사
	NNB	의존 명사
	NP	대명사
	NR	수사
	SWK	한글 자소
기호	SN	숫자
	SW	기타 기호
외국어	SL	외국어
	SH	한자

표 3은 본 논문에서 제안하는 토큰화 방법과 KoNLPy 라이브러리를 이용하여 명사를 추출한 결과이다. 제안하는 토큰화 방법에서는 '▶' 문자의 경우 한 글자 단어에 해당하여 추출되지 않는 것을 볼 수 있다. 또한 'C SINO' 문자열의 경우 알파벳의 A 대신 특수문자의 를 사용하였으며 알파벳의 O 대신 숫자의 0을 사용하여 'CASINO'(카지노)라는 단어를 표현하였다. 이러한 문자열의 경우, KoNLPy뿐 아니라 명사를 추출하는 일반적인 라이브러리에서는 명사로 인식되지 않아 추출되지 않는 것을 볼 수 있다. 또한 '10%', '500만'처럼 숫자와 기호 또는 글자가 더해져 의미를 갖는 경우에도 하나의 단어로 추출이 가능한 것을 볼 수 있다.

(표 3) 제안하는 토큰화 결과와 KoNLPy 명사 추출 라이브러리 결과 비교

(Table 3) Comparison of Our Tokenization Results with KoNLPy Noun Extraction Library Results

Method	Tokenization Result
None (원문)	▶ -C SINO▶ 5만제공+10%(500만)환급드림!jja86. c m무료거부0805426482
KoNLPy	환급, 드림, 무료, 거부
제안 방식	C SINO, 5만제공, 10%, 500만, 환급드림, jja86, c m, 무료거부, 0805426482

다음으로 토큰화가 완료된 문자열을 이용하여 각 카테고리 별 신조어 후보 사전들을 구축한다. 구축한 신조어 후보 사전은 내림차순으로 정렬하여 사용한다. 본 실험에서는 =0, =5로 설정하였다. 신조어 후보 사전의 토큰에 대해 가능한 많은 신조어를 감지하기 위해 값을 0으로 설정하였다. 또한 ConceptNet의 Web API를 이용하여 여러 단어들의 검색 결과를 살펴본 결과, '대한민국', '사랑', '여름'과 같은 단어의 경우 19~20개의 관련 단어가 검색되었지만, '대리', '대출'의 경우 각각 0개, 2개의 관련 단어만 검색되었다. 이를 통해 ConceptNet 검색 결과가 5미만일 경우 신조어로 감지하도록 설정하였다. 파라미터 와 는 환경에 따라 조정하여 사용한다.

(표 4) 신조어 감지 결과

(Table 4) Korean New Words Detection

Category	Korean New Words
도박	일주일간유료, 6년차널터해외본사, 봄햇살, 잔금⑤이용바람, NiCE, 첫사용,
금융	캐시모아, 중도상환수수료, 신용대출, 자격기준, 가상조회, 주택담보대출혜택
선거	선거운동정보, 군민, 시장, 여론조사, 지지, 시민, 출마,
대리운전	공감대리, 10퍼센트누적직립, 센터운영직립, 획식된, 강남룸사롱
비스팸	페이스북, 매너콜, 인증번호, 고객센터, 국외발신, 배송,

표 4는 카테고리에 대한 신조어 감지 결과의 일부이다. KoNLPy 라이브러리를 이용할 경우, '고객', '강남' 등의 일반적인 단어의 경우 추출이 가능하지만 표 4와 같이

‘봄햇살’, ‘VIP고객님’, ‘강남름샤롱’ 등 스팸에서 사용되는 단어는 추출되지 않는다. 반면 본 논문에서 제안하는 신조어 감지의 경우, 스팸에서 사용되는 단어뿐 아니라 ‘NiCE’와 같이 대소문자가 혼재되어 있거나 ‘강남름샤롱’과 같이 맞춤법과 다르게 표기되어 있는 문자열도 단어로 인식하여 신조어로 감지되는 것을 볼 수 있다.

4.3 신조어 적용 스팸 필터링

이러한 방법으로 감지된 신조어를 스팸 필터링에 적용할 경우, 성능 향상을 검증하기 위해 신조어를 사용하지 않은 스팸 필터링과 신조어를 적용한 스팸 필터링의 성능을 비교한다. 스팸 필터링에는 multi-class Naive Bayes (NB)를 사용한다.

기본 NB에 신조어를 적용하기 위해 신조어 발생 횟수를 활용하여 가중치를 계산하고, 이를 기본 NB 카테고리별 확률 값에 가중치를 더한 값으로 스팸을 필터링하는 방식을 사용하였다. 신조어 사전의 경우 카테고리 별로 구분되어 있기 때문에 가중치 역시 카테고리에서 신조어 발생 비율로 계산한다. 가중치 계산 시, 상위 신조어 발생 횟수와 하위 신조어 발생 횟수의 차이가 심하여 편차를 줄이기 위해 신조어 발생 횟수를 카테고리 내 전체 발생 횟수로 나눈 값에 루트 값을 사용하였다.

신조어를 적용한 NB는 두 단계로 진행된다. 먼저 기본 NB를 수행하여 각 카테고리에 속할 확률을 계산한다. 다음으로 각 스팸 문자에 신조어가 존재할 경우, 해당 신조어가 속한 카테고리의 확률에 계산한 가중치 값을 더해 주어 해당 카테고리에 속할 확률을 높여준다. 이 때, 확률 값에 가중치를 계속적으로 더해줄 경우 확률 값이 1 이상 될 수 있다. 이를 방지하기 위해 업데이트된 확률 값은 softmax 함수를 사용하여 모두 0~1 사이 값으로 정규화하며 정규화된 확률 값에서 최댓값을 갖는 카테고리로 재분류한다.

표 5를 보면 기본 스팸 필터링의 평균 F1 스코어는 0.84, 신조어 적용 스팸 필터링의 평균 F1 스코어는 0.90으로 신조어가 적용되었을 때 성능이 향상하는 것을 확인할 수 있다. 기본 스팸 필터링의 경우 ‘비스팸’ 카테고리의 재현율과 F1 스코어가 다른 카테고리에 비해 현저히 낮은 것을 확인할 수 있다. 이는 스팸이 아닌 모든 문자가 비스팸에 속하기 때문에 비스팸 문자의 패턴을 파악하기 어려워 재현율이 낮고 이로 인해 F1 스코어도 낮아지게 된다. 그러나 신조어가 적용된 스팸 필터링에서는 ‘비스팸’ 카테고리의 재현율과 F1 스코어가 눈에 띄게 향

(표 5) 기본 스팸 필터링과 신조어 적용 스팸 필터링의 성능 비교

(Table 5) Comparison of Basic Spam Filtering with Spam Filtering applied Korean New Words

Category		Precision	Recall	F1-score
기본 스팸 필터링	도박	0.87	0.94	0.90
	선거	0.89	0.91	0.90
	대리운전	0.98	0.87	0.92
	금융	0.81	0.90	0.86
	비스팸	0.88	0.52	0.66
	평균	0.88	0.82	0.84
신조어 적용 스팸 필터링	도박	0.94	0.89	0.92
	선거	0.92	0.94	0.93
	대리운전	0.93	0.95	0.94
	금융	0.88	0.93	0.91
	비스팸	0.80	0.85	0.82
	평균	0.89	0.91	0.90

상되었다. NB에서 비스팸으로 분류하지 않아도 문장에서 비스팸에 해당하는 신조어가 많이 등장할 경우 ‘비스팸’ 카테고리에 대한 가중치로 인해 확률 값이 증가하여 ‘비스팸’으로 재분류한다.

5. 결론 및 향후 연구

본 논문에서는 변화하는 스팸 문자에 대응할 수 있는 스팸 필터링 시스템의 구현과 지식 그래프에서 한국어를 비롯한 공통 언어의 경우 어휘 수가 매우 부족한 문제를 개선하기 위해 신조어 감지 매커니즘을 제안하였다. 더불어 콘텐츠 기반 스팸 필터링에서 기본적으로 사용되는 Naive Bayes에 신조어 감지 매커니즘을 적용한 실험을 통해 제안하는 방식의 성능을 확인하였다.

실험에서 구현한 토큰화 과정에서 더욱 정확한 품사 태깅이 가능하다면 양질의 단어 토큰으로 더욱 정교한 신조어 감지가 가능할 것으로 보인다. 또한 본 논문에서는 토큰을 추출하는 과정에서 명사, 기호, 외국어 세 가지 품사만을 고려하였으나 이후 URL, 전화번호 등 스팸의 추가적인 특성까지 고려하여 토큰화를 진행할 경우 스팸 필터링에 더욱 유의미한 신조어 추출이 가능할 것으로 보인다.

스팸 필터링이 발전함에 따라 스팸에서 사용되는 단어 역시 변화한다. 이 때, 컴퓨터에서 단어의 변화는 새로운 단어, 즉 신조어로 인식된다. 따라서 스팸 필터링이 변화하는 스팸 패턴에 대응하기 위해서는 스팸 문자에서 사

용되는 새로운 단어를 감지하는 것이 중요하다.

향후에는 본 논문에서 제안하는 신조어 감지 매커니즘을 대화형 챗봇에 적용하여, 대화에 신조어가 탐지되었을 때 자동으로 지식 그래프에 추가하고 그 의미를 이해할 수 있는 지능형 대화형 챗봇에 대해 연구할 예정이다.

참고문헌(Reference)

- [1] Y. E. Jo, "NewsLite", <http://www.newsgg.net/mobile/article.html?no=13807>
- [2] H. S. Ahn, "Safetimes", <http://www.safetimes.co.kr/news/articleView.html?idxno=76901>
- [3] Rojas-Galeano, Sergio A, "Revealing non-alphabetical guises of spam-trigger vocables", *Dyna*, Vol.80, No.182, pp.15-24, 2013. <http://ref.scielo.org/k4w22k>
- [4] Joe, In-Whee, and Hye-Taek Shim, "A SVM-based spam filtering system for short message service (SMS)", *The Journal of Korean Institute of Communications and Information Sciences*, Vol.34, No.9, pp.908-913, 2009. <http://www.koreascience.or.kr/article/JAKO200933063799701.page>
- [5] Kang, Seung-Shik, "A Normalization Method of Distorted Korean SMS Sentences for Spam Message Filtering", *KIPS Transactions on Software and Data Engineering*, Vol.3, No.7, pp.271-276, 2014. <https://doi.org/10.3745/KTSDE.2014.3.7.271>
- [6] Narayan, Akshay, and Prateek Saxena, "The curse of 140 characters: evaluating the efficacy of SMS spam detection on android", *Proceedings of the Third ACM workshop on Security and privacy in smartphones & mobile devices*. ACM, 2013. <https://doi.org/10.1145/2516760.2516772>
- [7] Wang, Zhen, et al, "Knowledge graph embedding by translating on hyperplanes", *Twenty-Eighth AAAI conference on artificial intelligence*, 2014. <https://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8531/8546>
- [8] Lin, Yankai, et al, "Learning entity and relation embeddings for knowledge graph completion", *Twenty-ninth AAAI conference on artificial intelligence*, 2015. <https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9571/9523>
- [9] Speer, Robert, Joshua Chin, and Catherine Havasi, "Conceptnet 5.5: An open multilingual graph of general knowledge", *Thirty-First AAAI Conference on Artificial Intelligence*, 2017. <https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14972/14051>
- [10] Speer, Robert, and Catherine Havasi, "Representing General Relational Knowledge in ConceptNet 5", *LREC*, 2012. http://lrec-conf.org/proceedings/lrec2012/pdf/1072_Paper.pdf
- [11] Karami, Amir, and Lina Zhou, "Improving static SMS spam detection by using new content-based features", *Twentieth Americas Conference on Information Systems*, Savannah, 2014. <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1205&context=amcis2014>
- [12] Han-Cheol Cho, and Geun-Sik Jo, "Spam-mail Filtering System Using Naive Bayesian Classifier and Message Rule", *Proceedings of the KISS conference, 한국정보과학회 학술발표논문집*, Vol.29, No.1B, pp.223-225, 2002. <http://www.dbpia.co.kr/pdf/pdfView.do?nodeId=NODE00612490>
- [13] Xiang, Yang, Morshed Chowdhury, and Shawkat Ali, "Filtering mobile spam by support vector machine", *CSITeA'04: Third International Conference on Computer Sciences, Software Engineering, Information Technology, E-Business and Applications*. International Society for Computers and Their Applications (ISCA), 2004. <http://hdl.handle.net/10536/DRO/DU:30005386>
- [14] Gómez Hidalgo, José María, et al, "Content based SMS spam filtering", *Proceedings of the 2006 ACM symposium on Document engineering*. ACM, 2006. <https://doi.org/10.1145/1166160.1166191>
- [15] Duan, Longzhen, Nan Li, and Longjun Huang, "A new spam short message classification", *2009 First International Workshop on Education Technology and Computer Science*, Vol.2, pp.168-171, 2009. <https://doi.org/10.1109/ETCS.2009.299>
- [16] Liu, Wuying, and Ting Wang, "Index-based online text

- classification for sms spam filtering”, Journal of Computers, Vol.5, No.6, pp.844-851, 2010.
<https://doi.org/10.4304/jcp.5.6.844-851>
- [17] Huang, Jie, Bei Huang, and Wenjing Pu, “A Bayesian approach for text filter on 3G network”, 2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM). IEEE, 2010. <https://doi.org/10.1109/WICOM.2010.5601282>
- [18] Roy, Pradeep Kumar, Jyoti Prakash Singh, and Snehasish Banerjee. “Deep learning to filter SMS Spam.”, Future Generation Computer Systems, Vol.102, pp.524-533, 2020.
<https://doi.org/10.1016/j.future.2019.09.001>
- [19] Dhavale, Sunita. “C-ASFT: Convolutional Neural Networks-Based Anti-spam Filtering Technique.”, Proceeding of International Conference on Computational Science and Applications. Springer, Singapore, pp.49-55, 2020.
https://doi.org/10.1007/978-981-15-0790-8_6
- [20] Sharmin, Tazmina, et al. “Convolutional neural networks for image spam detection.”, Information Security Journal: A Global Perspective pp.1-15, 2020.
<https://doi.org/10.1080/19393555.2020.1722867>
- [21] Ezpeleta, Enaitz, et al. “Novel email spam detection method using sentiment analysis and personality recognition.”, Logic Journal of the IGPL, Vol.28, No.1, pp.83-94, 2020.
<https://doi.org/10.1093/jigpal/jzz073>
- [22] Venkatraman, S., B. Surendiran, and P. Arun Raj Kumar. “Spam e-mail classification for the internet of things environment using semantic similarity approach.”, The Journal of Supercomputing, Vol.76, No.2, pp.756-776, 2020.
<https://doi.org/10.1007/s11227-019-02913-7>

● 저 자 소 개 ●



김 지 혜(Ji-hye Kim)

2011년 가천대학교 소프트웨어설계경영학과 (공학사)

2019년 가천대학교 대학원 소프트웨어학과 (석사과정)

관심분야 : Big Data, Socialmedia Mining, Machine Learning, Deep Learning, etc.

E-mail : kimjihae28@gachon.ac.kr



정 옥 란(Ok-ran Jeong)

2005년 이화여자대학교 대학원 컴퓨터공학과 (공학박사)

2006년 서울대학교 컴퓨터공학부 (박사 후 연구원)

2007년 Univ. of Illinois at Urbana-Champaign (박사 후 연구원)

2008년 성균관대학교 정보통신학부 (연구교수)

2009~현재 가천대학교 소프트웨어학과 (부교수)

관심분야 : Big Data, Socialmedia Mining, Machine Learning, etc.

E-mail : orjeong@gachon.ac.kr