특집 02

의약품 부작용 예측을 위한 빅데이터 분석 기술 동향

김현희 (동덕여자대학교)

목 차

1. 서 론

의약품은 여러 차례에 걸친 임상 시험을 통과하고 안정성을 평가받은 후 시판이 허가되지만, 임상 시험 단계에서 발견하지 못했던 중대한 부작용이 나타날 가능성이 항상 존재한다. 임신 초기 입덧에 효과적인 항구토제로 수년간 시판되었던 탈리도마이드는 후에 태아에서 사지결손증을 유발한다는 사실이 밝혀져 시판이 중지 되었으나, 이미 전세계 46개국에서 12,000명 이상의 기형아 출산을 야기하였다. 이 때 태어난 아이 중 40 %가 그해에 사망하였고 현재 생존자는 약 5,000명 정도로 추정된다[1]. 우리나라에서도 감기약 치료제로 널리 사용되었던 페닐프로판올아민 성분 제제가 출혈성 뇌졸중을 야기하는 것으로 밝혀져 판매 중지 되는 등 최근 10년간 국내외에서 의약품이 시판된 이후에 안전성 혹은 유효성 문제로 허가 취소되거나 판매 중지된 사례는 수십 건에 달한다[2].

의약품 부작용 (Side Effect)이란 의약품 등을 정상적인 용법에 따라 투여한 경우 발생하는 모든 의도되지 않은 효과를 말하며, 이 중에서 바람직하지 않은 징후, 증상 또는 질병을 이상 사례 (Adverse Event) 라고 한다[3]. 이러한 이상 사례의 원인은 허가 시점에서 얻을 수 있는 의약품 정보의 한계에서 기인한다. 즉, 시판 전 임상 시험 시 소아, 임산부 및 노인에 대한 시험이 제한되고 있으며, 임상 시험 후 일정 기간 동안 부작용 정보를 추적하므로 장기간이 지난 뒤 나타나는 부작용에 대한 발견이 어렵다. 또한 환자 개개인의 질환이나 병용되는 약물에 대한 고려되지 않는 경우가 많다. 따라서 의약품의 시판 후에도 부작용을 파악하고 관리하는 것이 필수적이다.

이와 같이 의약품 관련 문제의 탐지, 평가, 해석 및 예방에 관한 과학적 연구 및 활동을 약물

감시 (Pharmacovigilance) 라고 하는데[3], 약물 감시를 위해 자발적 부작용 보고 시스템이 구축되어 운영되고 있다. 미국 FDA에서 운영하는 자발적 부작용 신고 시스템 (Federal Drug Administration's Adverse Event Reporting System, FAERS) 의 경우 매년 60만건 이상의 부작용 신고가 접수되고 있으며, 세계 보건 기구 (World Health Organization)에서는 1968년 국제 약물 감시 체계를 구축하였고, 스웨덴의 웁살라 모니터링센터 (Uppsala Monitoring Center)에서는 1978년부터 전세계로부터 의약품 부작용 사례를 보고받고 있다[4]. 한국의 경우 한국 의약품 안전관리원에서 의약품 부작용 보고 시스템을 구축하고 데이터베이스를 관리 및 정보 제공을 담당하고 있다. 그러나 자발적 부작용 보고 자료는 의사나 약사, 간호사 등 의약 전문가나 제조회사에서 자발적으로 부작용 정보를 제공하는 것이므로 과소 보고의 제약점을 안고 있으며, 보고된 사례도 보고자에 따라 정보의 질이 현격히 다르다는 단점이 있다.

이와 같은 문제점을 극복하기 위해서 최근에는 부작용 보고 데이터베이스뿐만이 아니라 다양한 목적으로 구축된 데이터베이스들을 통합하여 정형화된 빅데이터를 구축하고 이를 약물 감시에 활용하고자 하는 연구가 활발히 이루어지고 있다. 특히 자발적으로 보고된 이상 사례 중에서 해당 의약품과의 인과관계를 배제할 수 없는 경우를 약물 이상 반응 (Adverse Drug Reactions)이라고 하는데, 빅데이터 분석을 통해서 약물 이상 반응을 탐지하고 약물과 약물 이상 반응과의 인과 관계인 실마리 정보를 찾아낸다면, 보다 신속한 의약품 부작용 관리가 이루어질 수 있다. 미국의 경우 하버드 대학을 협연 센터로 지정하고 전자 건강보험 청구자료, 입원 및 외래환자의 의무 기록, 환자 등록 자료 등을 통합하여 분산 데이터 분석 체계를 구축하였고, 유럽에서는 유럽의약품청 주도로 European Network of Centers for Pharmacoepidemiollogy and Pharmacovigilance (ENCePP)를 운영하여 유해성 조기 파악 방법론을 개발하고 있다. 국내에서도 병원전자의무기록자료, 건강보험심사평가원의 요양급여 청구자료 등을 연계하여 의약품과 부작용간 인과성 평가를 실시하였다[5].

또한 환자들이 소셜 네트워크에 올린 글이나 웹상의 게시물 등 비정형 텍스트 데이터를 활용하여 약물 이상 반응을 탐지하는 연구도 활발히 이루어지고 있다. 환자들의 소셜 네트워크 서비스인 patientslikeme[6]의 경우, 같은 질환을 가진 환자들끼리 증상, 부작용, 처방 기록 등의 정보를 주고받는 사이트로 기존의 데이터베이스에서 간과한 환자들의 의견이 고스란히 반영된 질병 정보 데이터를 제공하고 있다. 이와 같은 소셜 미디어 빅데이터의 활용은 임상 시험에 비하여 다양한 연령층 및 질환군을 포함할 수 있고, 의약품의 장기 사용에 의한 부작용 정보를 발견할 수 있다는 장점을 갖는다. 뿐만 아니라 시판 이후 부작용이 발견되고 평가를 거쳐 시판이 철회되기까지 비교적 많은 시간이 소요되므로 빅데이터 분석을 통한 부작용 예측은 그 시간을 단축시키는데 중요한 역할을 할 수 있다.

본 고에서는 제 2장에서 데이터 마이닝을 이용하여 부작용 보고 자료로부터 실마리 정보를 찾기 위한 분석 기법들을 알아보고, 제 3장에서 비정형 텍스트 분석을 통한 부작용 보고자료 분류 및 약물 이상 반응 탐지를 위한 소셜 미디어 분석에 활용된 기계 학습 기법들을 살펴본다. 제 4장에서 딥러닝을 활용한 디지털 약물 감시에 대해 소개한 다음, 제 5장에서 결론을 맺도록 한다.

2. 데이터 마이닝 기반 실마리 정보 검색 (signal detection) 기법

현재까지 가장 많이 알려진 부작용 실마리 정보 검색 기법은 미국 FAERS 데이터베이스에 데이터 마이닝 알고리즘을 적용하여 정량적인 방법으로 실마리 정보를 찾는 것이다[7,8]. 이중에서도 보고분율비 (Proportional reporting ratios, PRRs)와 보고오즈비 (Reporting odds ratios, RORs)가 가장 일반적이며 해석하기 쉬운 실마리 정보 검색 기법으로 널리 사용되고 있다. 표 1에서 보는 바와 같이 데이터베이스에 보고된 전체 보고건을 n 이라 하고, 약물 i가 갖는 부작용 j에 대한 보고건을 n_{ij} 이라고 하자. 이때 n_i는 약물 i 에 대한 보고건이고, n_j는 부작용 j에 대한 보고건이다[9].

보고분율비란 특정 약물 보고건의 특정 부작용 분율을 다른 약물 보고건의 부작용 분율로 나눈값을 말하며 다음 식과 같이 정의된다.

$$PRP = \frac{n_{ij/} n_i}{(n_j - n_{ij})/(n - n_i)}$$

보고오즈비란 특정 약물에 노출된 환자에서 발생한 특정 부작용 발생 오즈(odds)를 다른 약물에 대한 특정 부작용의 발생 오즈(odds)로 나눈 것으로 다음 식과 같이 정의된다.

$$ROR = \frac{n_{ij/} (n_j - n_{ij})}{(n_i - n_{ij})/(n - n_i - n_j + n_{ij})}$$

이밖에도 약물과 부작용간의 상관관계를 베이지안 방법을 활용하여 측정하는 Bayesian confidence propagation neural network (BCPNN) 방법[10]도 실마리지표 산출에 사용되고 있으며, 다양한 통계 기반의 데이터 마이닝 알고리즘들이 개발되고 있다. 이와 같은 데이터 마이닝 기반의 실마리 정보 검색 기법은 임상 시험이 갖는 여러 가지 제약점들을 극복하고 약물과 약물이상반응과의 인과 관계를 보다 신속하게 발견할 수 있다. 다만, 자발적 부작용 보고 제도는 과소 보고 문제를 항상 포함하고 있으므로 데이터 마이닝을 활용한 실마리 정보는 인과 관계의 가능성을 찾아주는 도구로서 보아야 할 것이다.

3. 기계 학습 기반 비정형 텍스트 분석과 활용

보다 최근에는 정형화된 의료 정보 빅데이터 뿐만이 아니라 비정형 텍스트 데이터에 대한 관심이 높아지고 있다. 바이오의약학 문헌은 약물간의 상호작용에서 오는 부작용 예측에 활용되고 있으며, 트위터와 같은 소셜 미디어도 환자들의 직접적인 목소리를 반영하므로 적극적으로 빅데이터 분석에 활용되고 있다. 텍스트 데이터

〈표 1〉 2 × 2 분할표

	부작용 j에 대한 보고건	부작용 j를 제외한 보고건	전체 보고건
약물 i에 대한 보고건	n_{ij}	$n_i - n_{ij}$	n_i
약물 i를 제외한 보고건	$n_j - n_{ij}$	$n - n_i - n_j + n_{ij}$	$n - n_i$
전체 보고건	n_j	$n - n_j$	n

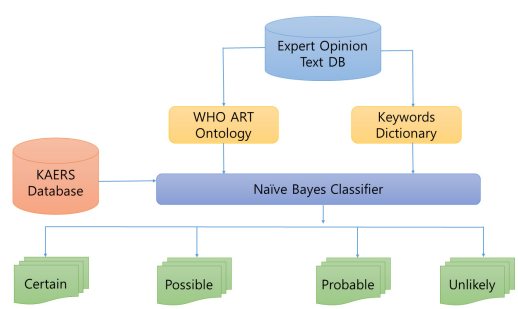
의 경우는 직접적으로 부작용 예측에 활용되기보다는 부작용 보고 자료를 기계 학습을 통해서 자동 분류하고, 소셜 미디어 데이터를 분석하여 약물 이상 반응에 대해 언급한 메시지를 분류하는 등 의약품 부작용을 파악하고 관리하기 위한 보조적인 툴로서 활용되고 있다.

본 장에서는 먼저 텍스트 데이터를 활용한 의약품 부작용 보고자료의 자동 분류 기법을 소개하고, 다음으로 소셜 미디어를 분석하여 약물 이상 반응을 탐지하기 위한 기계 학습 기법을 소개한다.

3.1 텍스트 데이터를 활용한 의약품 부작용 보고자료 분류

국내에서는 한국의약품안전관리원의 의약품이상사례보고시스템 (The Korea Adverse Event Reporting System, KAERS)을 통해서 자발적 보고자료를 등록하고 있다. 이때, 의사, 약사, 혹은 제약회사와 같은 보고자는 보고자료와 실제 부작용과의 인과성을 평가하고 평가에 대한 의견을 자연어로 기술하고 있다. 전문가 의견 텍스트 데이터를 가진 보고건수는 1989년에서 2015년 6월까지 약 90,552건으로 아직까지 전문가 의견 텍스트 데이터가 의약품 부작용 연구를 위해 활용된 적이 없다[11]. 일반적으로 부작용 보고자료는 인과 관계 평가기준으로 정의된 가이드라인에 따라서 확실함(certain), 상당히 확실함(probable), 가능함(possible), 가능성 적음(unlikely), 평가곤란(unclassified), 그리고 평가불가(unassessable)의 7개 카테고리로 나누어진다. 전문가 텍스트를 활용하여 보고자료를 인과관계에 따라 4개의 범주로 자동분류하고, 인과관계가 확실한 보고자료들만 미리 선별하여 부작용 보고 데이터베이스 활용 시 데이터의 질을 향상시키고자 하였다[12].

그림 1은 전문가 의견 텍스트 데이터베이스를 활용한 보고자료 자동 분류 구조를 나타낸다. 자동 분류를 위해서 나이브 베이즈 알고리즘을 사용하였으며, 학습 단계에서 각 카테고리에 해당하는 단어들을 효율적으로 학습시키기 위해 tf-idf 가중치 기반의 단어 사전을 구축하였다. 단어 사전은 전문가 의견 문서를 텍스트 마이닝하여 주요 단어를 추출하여 반자동적으로 구축하였다. 또한 보고자가 서술한 전문가 의견이 국영문 혼합문으로 되어 있으므로 이를 처리하기 위해서 WHO Adverse Reaction Terminology (WHO-ART) 온톨로지를 구축하였다. WHO-ART 코드[13]는 WHO에서 정의한 약물 부작용에 관한 표준 용어 정의로서, 국문 용어를 영문으로 통합하기 위해 사용하였다. 예를 들면, 국문으로 사용된 “구토”라는 단어는 WHO-ART 온톨로지에 의해 “vomiting”이라는 표준 용어로 변환되고 이를 통해 분류 정확도를 향상시켰다.



(그림 1) 텍스트 데이터를 활용한 부작용 보고자료 자동 분류 구조

3.2 약물 이상 반응 탐지를 위한 소셜 미디어 분석

임상 시험은 임상 시험 대상자의 연령, 임상 시험 기간, 규모 등 제약을 가진 환경에서 실시

되는 반면 소셜 미디어를 활용하면 보다 다양한 환자들의 실시간 데이터를 분석하여 부작용 예측에 활용할 수 있다는 장점이 있다. 특히 트위터 게시물에 언급된 약물 이상 반응의 경우 FAERS에 보고된 사례와 통계적으로 유의한 상관 관계가 있음이 입증되었다[14]. 트위터 게시물을 분석하여 약물 이상 반응을 탐지하는 연구의 대부분은 해당 트위터 게시물이 약물 이상 반응에 해당하는 게시물인지 그렇지 않은지를 분류하기 위해 지도 학습을 활용한이진 분류가 주를 이루고 있다. 트위터 게시물에 사용되는 용어들은 약물 이상 반응을 나타내는 표준화된 용어들과 상이하므로 국제 분류 체계인 Medical Dictionary for Regulatory Activities (MedDRA)에서 정의한 용어들로의 매핑이 필요하다. 뿐만 아니라 잘못된 철자나 인터넷 용어 등 증상을 나타내는 다양한 표현들을 표준 용어를 변환하는데 많은 전처리 작업이 요구되며, 용어 사전이나 온톨로지를 사용하여 용어의 표준화를 실시하고 있다.

[15]에서는 나이브 베이즈, 서포트 벡터 머신, 그리고 최대 엔트로피 기반 분류기를 활용하여 트위터 메시지를 이진 분류하였다. 성능을 향상시키기 위해서 약물 이상 반응에 관련된 어휘들을 정의하였으며, tf-idf 가중치를 사용하여 유사어 집합을 정의된 어휘들에 포함시켰다. 또한 토픽 모델링을 적용하여 특정 토픽에 해당하는 키워드를 찾아내어 어휘 정의에 활용하였다. 실험결과 서포트 벡터 머신이 다른 분류기에 비하여 가장 뛰어난 성능을 보임을 확인하였으며, 특히, 지도학습 기반의 분류기의 경우 토픽 모델링이나 감성 분석, 그리고 다중 코퍼스 활용 등을 통한 약물 이상 반응과 관련된 용어의 추출이 텍스트 분류의 성능 향상에 크게 기여했다는 것을 알 수 있다.

4. 딥러닝 기반 디지털 약물 감시

FAERS에 의해 보고되는 부작용 사례는 실제 사례의 1-13 % 정도로 추정되고 있으며, 보고까지 시간이 걸리는 경우가 많고, 보고 내용도 내원 환자의 경우로 치우친 경우가 대부분이다[16]. 이러한 점에서 소셜 미디어를 통한 부작용 보고는 다양한 연령대를 포함하며, 신속하게 보고되므로 기존의 자발적 의약품 부작용 보고 제도를 보완할 수 있는 대안으로 떠오르고 있다. 소셜 미디어를 활용한 약물 이상 반응 탐지를 디지털 약물 감시라고 하며 웹 페이지, 환자들의 소셜 네트워크 서비스, 트위터 등으로부터 약물이나 건강 상태 혹은부작용에 대한 텍스트를 수집하여 다양한 분석을 시도하고 있다. 대표적으로 활용되고 있는 소셜 미디어인 트위터 분석의 경우, 약물과 증상 및 부작용에 대한 용어 정의 및 표준 용어로의 매핑 작업이 큰 부분을 차지하며 용어 사전의 활용이 분석 성능을 좌우한다. 이러한 문제점을 보완하기 위해서 최근에 트위터 분석을 위해 딥러닝을 활용하여 트위터 메시지로부터 약물 이상 반응을 탐지하고자 하는 연구가 이루어지고 있다[16].

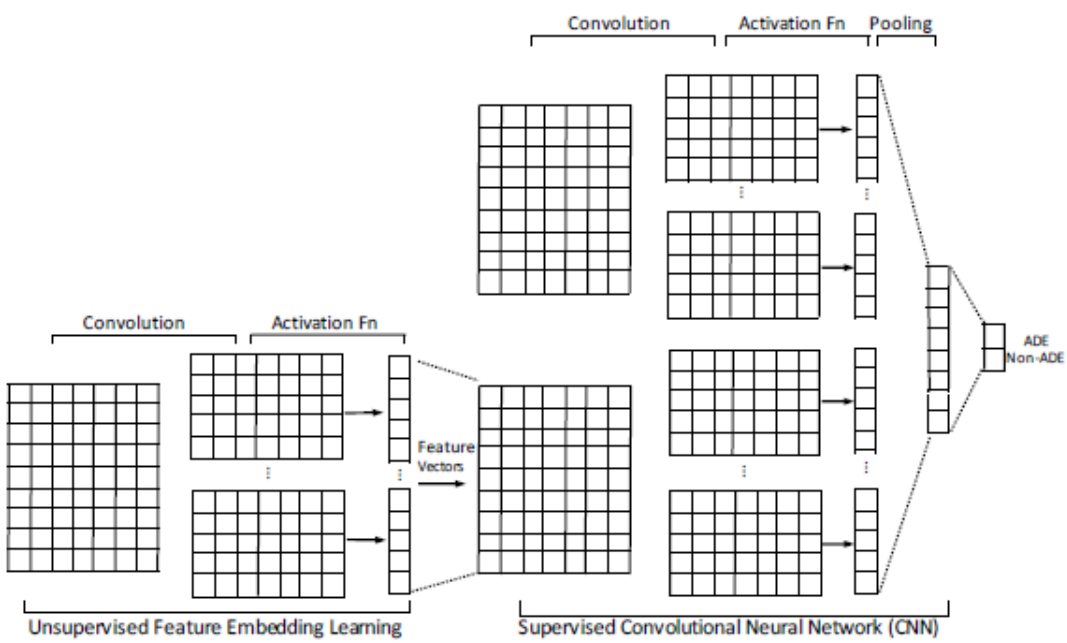
딥러닝 모델은 전통적인 기계 학습 알고리즘과 달리 도메인 지식을 활용하지 않고 데이터를 학습하여 특징 추출이 이루어지므로 성공적으로 대량의 텍스트 분류에 활용되고 있다. 따라서 트위터 메시지로부터 약물 이상 반응에 해당하는 메시지를 분류하는 작업 역시 딥러닝이 효과적으로 활용될 수 있다. 그러나 딥러닝이 좋은 성능을 발휘하려면 명확하게 분류된 대량의 텍스트가 필요한데, 실제 트위터 메시지 중에서 약물 이상 반응에 대해 언급한 메시지는 턱없이 부족하다. 이러한 트위터 데이터의 제약점을 극복하

고자 준지도 학습 기반 컨볼루션 신경망 모델이 활용되었다.

그림 2는 컨볼루션 신경망을 이용하여 트위터 메시지로부터 약물 이상 반응을 검출하는 방법을 나타낸다[17]. 이 방법은 크게 두 단계로 나뉘는데, 먼저 비지도 학습 단계를 통해서 트위터에서 사용되는 의학 용어, 약물명, 건강 상태 및 약물과 건강상태를 나타내는 구문들을 학습시킨다. 다음으로 학습한 구문 임베딩과 레이블을 갖는 트위터 메시지들을 가지고 약물 이상 반은 탐지 분류기를 위해 다시 학습시킨다. 약물에 관한 메시지, 건강 조건에 관한 메시지, 약물 조건에 관한 메시지 등 다양한 모델을 생성하고 실험을 실시한 결과 전통적인 기계 학습 분류기를 사용한 경우보다 높은 분류 성능을 보였다.

5. 결 론

임상 시험이 갖는 현실적인 제약 사항 때문에 의약품에 대한 시판 후 감시 및 관리는 필수적인 요인이다. 자발적 부작용 보고 제도를 통해서 부작용 자료를 수집하고 분석하여 정보를 제공하고 있으나, 이는 실제 일어나는 부작용 사례에 비해서 매우 작은 수의 보고 자료를 기반으로 한다. 따라서 전자 건강 기록 및 보험 청구 데이터베이스 등과 같이 다양한 목적으로 구축된 데이터베이스들을 빅데이터로 통합하여 보다 유용한 정보를 제공하고자 하는 노력이 이루어지고 있다. 또한 바이오 의약학 전문 서적이나 논문이나 전문가 의견 문서등과 같은 비정형 텍스트를 분석하여 부작용 실마리 정보를 찾는데 도움을 주고자 하는 연구도 활발하다. 보다 최근에는 실시간으로 환자들의 건강 상태를 반영하는 소셜 미디어 데이터를 활용한 디지털 약물 감시도 점차



(그림 2) Semi-Supervised CNN[17]

로 중요성을 더하고 있다.

전통적인 통계적 데이터 마이닝 기법은 정형화된 빅데이터로부터 실마리 정보를 찾는데 널리 활용되어 왔다. 비정형 텍스트 데이터의 경우는 지도 학습 기반의 분류 알고리즘을 이용하여 약물 이상 반응과 관련된 텍스트를 분류하는 연구들이 진행되고 있다. 특히, 트위터와 같은 소셜 미디어의 경우는 전통적인 기계 학습 분류기보다 딥러닝을 활용하는 것이 분류 성능을 향상시킴을 알 수 있다.

현재 빅데이터는 의약학 분야에서 의약품 부작용의 예측 외에도 신약 개발을 위한 후보 물질 선정이나 개인별 맞춤 약물 치료제 개발 등에 활용되고 있다. 이러한 빅데이터 활용 분야 전반에 걸쳐 많은 투자와 연구 및 개발이 필요한 시점이다.

참 고 문 헌

[1] WG. McBride, Thalidomide and congential abnormalities. Lancet, Vol 2, pp. 1358, 1961.

[2] D. Choi, M. Choi, and A. Ko, Current status of pharmaceutial safety management in Korea, J. Korean Med. Assoc. Vol. 55, No. 9, pp. 827-834, 2012.

[3] 한국의약품안전관리원, http://www.drugsafe.or.kr/

[4] NK Choi, J. Lee and BJ Park, Recent international initiatives of drug safety management, J. Korean Med. Assoc. Vol. 55, No. 9, pp. 819-826, 2012.

[5] BJ Park, Application of big data for public health, J. Korean Med. Assoc. Vol. 57, No. 5, pp. 383-385, 2014.

[6] patientslikeme, http://www.patientslikeme.com/

[7] A. Bate and SJ Evans, Quantitative signal detection using spontaneous ADR reporting, Pharmacoepidemiol Drug Saf. Vol. 18, pp. 427-436, 2009.

[8] T. Tamura, T. Sakaeda, K. Kadoyama, et al. Aspirin- and clopidogrel-associated bleeding complications: Data Mining of the public version of the FDA Adverse Event Reporting System, Int. J. Med. Sci, Vol. 9, pp. 441-446, 2012.

[9] K. Sarvnaz et al., Text and Data Mining Techniques in Adverse Drug Reaction Detection, ACM Computing Surveys, Vol. 47, No. 4, pp. 56:1 - 56: 39, 2015.

[10] T. Sakaeda, A. Tamon, K. Kadoyama, and Y. Okuno, Data Mining of the Public Version of the FDA Adverse Event Reporting System, Int. J. of Med. Sci., Vol. 10, pp. 796-803.

[11] H. Kim and KY Rhew, Analysis of Adverse Drug Reaction Reports Using Text Mining, Korean J. Clin. Pharm, Vol. 27, No. 4, pp. 221-227, 2017.

[12] H. Kim and KY Rhew, A Machine Learning Approach to Classification of Case Reports on Adverse Drug Reactions using Text Mining of Expert Opinions, Lecture Notes in Electronic Engineering, Vol. 474, pp. 1072-1077, 2018.

[13] KH Lim, et al, Comparison of WHO-ART Versus MedDRA, Internationally Standardized Terminology of Adverse Drug Reaction Classification, Korean. J. Cli. Pharm. Vol. 17, No. 1, pp. 46-52, 2007.

[14] CC Freifeld, JS Brownstein, CM. Menone, et al., Digital drug safety surveilance: monitoring pharmaceutical products in Twitter, Drug Saf., Vol. 37, No. 5, pp. 343-350, 2014.

[15] A. Sarker and G. Gonzalez, Portable automatic text classification for adverse drug reaction detection via multi-corpus training, Journal of Biomedical Informatics, Vol. 53, pp. 196-207, 2015.

[16] A. Cocos, A. G. Fiks, and A. J. Masino, Deep

learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts, J. of the American Medical Inoformatics Association, Vol. 24, No. 4, pp. 813-821, 2017.

[17] K. Lee, et al, Adverse Drug Event Detection in Tweets with Semi-Supervised Convolutional Neural Networks, In Proc. of International World Wide Web Conference, Perth, Australia, pp. 705-714, 2017.

저 자 약 력

김 현 희

이메일 : heekim@dongduk.ac.kr

- 1996년 이화여자대학교 컴퓨터학과 (학사)
- 1998년 이화여자대학교 컴퓨터학과 (석사)
- 2005년 이화여자대학교 컴퓨터학과 (박사)
- 2005년~2006년 LG 전자 디지털 미디어 연구소 / 선임연구원
- 2006년~현재 동덕여자대학교 정보통계학과 부교수
- 관심분야 : 기계학습, 빅데이터 분석, 의약품 부작용 예측