정 보 보 호 학 회 지
제 29 권 제3호, 2019. 6

피싱 웹사이트 URL의 수준별 특징 모델링을 위한 컨볼루션 신경망과 게이트 순환신경망의 퓨전 신경망

부 석 준*, 김 혜 정**

요 약

폭발적으로 성장하는 소셜 미디어 서비스로 인해 개인간의 연결이 강화된 환경에서는 URL로써 전파되는 피싱 공격의 위험성이 크게 강조된다. 최근 텍스트 분류 및 모델링 분야에서 그 성능을 입증받은 딥러닝 알고리즘은 피싱 URL의 구문적, 의미적 특징을 각각 모델링하기에 적절하지만, 기존에 사용하는 규칙 기반 앙상블 방법으로는 문자와 단어로부터 추출되는 특징간의 비선형적인 관계를 효과적으로 융합하는데 한계가 있다. 본 논문에서는 피싱 URL의 구문적, 의미적 특징을 체계적으로 융합하기 위한 컨볼루션 신경망 기반의 퓨전 신경망을 제안하고 기계학습 방법 중 최고의 분류정확도 (0.9804)를 달성하였다. 학습 및 테스트 데이터셋으로 45,000건의 정상 URL과 15,000건의 피싱 URL을 수집하였고, 정량적 검증으로 10겹 교차검증과 ROC커브, 정성적 검증으로 오분류 케이스와 딥러닝 내부 파라미터를 시각화하여 분석하였다.

Ⅰ. 서 론

다양한 형태의 보안 위협 및 공격으로부터 사용자의 개인정보 및 컴퓨팅 리소스를 방어하는 네트워크 보안 기술은 정책과 방법의 두 가지 요소로 정의한다. 정책 위반 및 악의적인 활동으로부터 네트워크와 개인 사이버 자산을 보호하려는 다양한 방법이 제안되었고, 시스템이나 네트워크 관리자가 주기적으로 방어 메커니즘을 실행함으로써 능동적 공격을 탐지하고 대처하는 연구가 선행되었다 [1]. 반면에 시스템이 아닌 사용자를 속임으로써 사용자가 입력한 모든 정보를 탈취하는 피싱 공격에 대한 메커니즘은 상대적으로 부족하다.

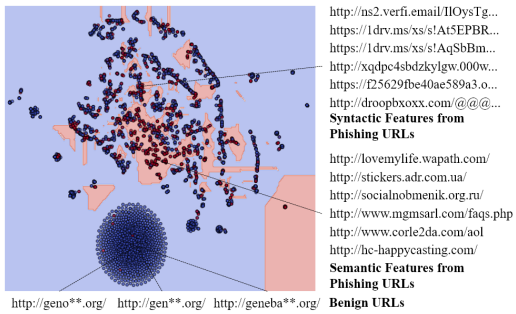
피싱 공격은 URL을 통해 사용자가 감염된 웹사이트로 이동하도록 유도하는 수동적 공격의 일종으로서, 해당 웹사이트에 입력되는 모든 정보는 공격자에게 전달되기에 치명적인 경제적 손실을 유발할 수 있다. 특히 개인간의 연결이 강화된 환경에서는 URL로써 전파되는 특징에 따라 그 위험성이 크게 강조된다 [2].

피싱 공격의 악성 URL의 탐지를 위해 제안된 여러 보안 시스템에서는 주로 피싱 데이터베이스를 활용하는 규칙 기반의 탐지가 주로 연구되었다 [3]. 그러나 URL의 발행이 상대적으로 간단한 웹 어플리케이션의 특성에 따라 피싱 URL은 매번 새로운 공격 인스턴스가 관측되는 Zero-day 공격의 특성을 지니고 있기 때문에 기관측된 데이터베이스와 규칙으로 대처할 수 없다 [4].

최근 여러 분야에서 데이터에 기반한 복잡한 비선형적 사상함수로써 활용되며 그 성능을 입증받은 딥러닝 알고리즘은 텍스트 데이터로부터 분류에 효과적인 특징을 자동으로 찾아낸다는 점에서 유용성을 입증받았다 [5]. 특히 입력된 텍스트 특징을 수준별로 모델링하기 위한 컨볼루션 신경망과 순환 신경망을 앙상블하는 시도에서는 악성코드의 분류 분야에 딥러닝 알고리즘 앙상블을 도입하여 유의미한 성능 향상을 보였다 [6]. 그림 1에서는 수준별 텍스트모델링이 필수적인 피싱 URL 데이터를 t-SNE 차원축소 방법으로 시각화하였다.

그림 1의 파란색과 붉은색 점은 각각 정상 URL과 피싱 URL 인스턴스를 나타낸다. URL을 구성하는 문자 조합의 유사성에 기반하여 유클리디안 거리가 결정되었고, 하단에는 주로 짧고 정규적인 정상 URL의 군

이 논문은 2019년도 정부(교육부)의 재원으로 한국 연구재단의 지원을 받아 수행된 기본연구지원사업(NO.2019B-010)의 연구수행으로 인한 결과물임을 밝힙니다.
* 연세대학교 컴퓨터과학과 (sjbuhan@yonsei.ac.kr)
** 경일대학교 사이버보안학과 (hjkim325@kiu.kr)



〔그림 1〕 피싱 URL 특징의 자연어 수준별 모델링과 그 융합방법의 필요성.

집이 형성되었다. 반면에 중앙에는 하위 도메인으로 인해 정상과 피싱 URL의 구분이 어려운 인스턴스들이 복잡하게 뒤섞여 있는데, 그중에서 피싱 URL들에 대해 구문적인 특징과 의미적인 특징으로 각각 구분하여 표기하였다.

피싱 URL은 주로 의미없는 문자의 시퀀스로 이루어져 구문적인 특징을 보이는 경우와 잘 알려진 단어의 시퀀스로 이루어지는 의미론적인 특징으로 구별할 수 있다. 기존의 딥러닝 앙상블에서는 구문/의미적인 컨볼루션 그리고 순환 신경망의 출력값을 로그-스케일에서 평균취하는 단순한 규칙을 사용하였으나, 그림 1에서 보이는 구문/의미적인 특징의 복잡한 비선형적인 관계를 효과적으로 모델링할 수 없는 한계를 보인다.

따라서 본 논문에서는 피싱 URL의 구문적인 특징과 의미적인 특징을 체계적으로 융합하기 위한 컨볼루션 신경망 기반의 퓨전 신경망을 제안한다. 컨볼루션 기반 퓨전 신경망은 데이터로부터 학습 가능한 필터 연산을 사용하여 피싱 URL 분류에 유효한 특징을 추출하는 딥러닝 알고리즘으로, 기존 컨볼루션 그리고 순환신경망으로부터 추출되는 URL의 분산 표현으로부터 최적의 앙상블 규칙을 학습하기에 적절하다. 제안하는 방법이 기존 딥러닝 알고리즘을 포함한 기계학습 방법 중에서 최고 분류 정확도(0.9803)를 달성함을 10겹 교차검증하였고, 통계적 유의미성을 카이-제곱 평가로 검증함으로써 피싱 URL 분류 분야에서 수준별 모델링과 체계적인 융합방법의 필요성을 보였다.

〔표 1〕 기계학습 기반 피싱 URL 분류 관련 연구

Author	Feature Extraction	Modeling
P. Prakash [7]	Lexical	Matching Rule
J.Ma [8]	BOW	Naive Bayes
A. Le [9]	BOW	SVM
R. Verma [10]	Lexical	Random Forest
A. Bahnsen [11]	W2V	LSTM
J. Zhao [12]	W2V	GRU
A. Anand [13]	Lexical	DCGAN
W. Yang [14]	W2V	CNN-LSTM

Ⅱ. 관련 연구

피싱 URL 분류 연구는 2010년도 초반까지 주로 연구된 블랙리스트 기반의 탐지, 전통적인 기계학습 기반의 텍스트로부터 추출한 어휘의 모델링, 그리고 최신의 딥러닝 알고리즘 기반의 텍스트 특징 추출 세가지 분야로 구분할 수 있다. 표 1에서 피싱 URL 분류를 위한 최근 10년간의 연구동향을 요약하였다.

P. Prakash et al. 에서는 텍스트로부터 추출할 수 있는 어휘적인 특징을 전문가의 규칙에 기반하여 추출하고, 알려진 피싱 URL에 대한 블랙리스트를 작성하여 단순한 비교 알고리즘을 사용하여 새로운 피싱 URL을 탐지하는 시스템을 제안하였다 [7]. 그러나 이 방법은 새로운 피싱 URL을 탐지하기에는 일반화 성능의 측면에서 한계가 있었다. J. Ma et al. 에서는 URL 에 포함된 단어 조합에 대해 Naive Bayes 를 포함하는 기본적인 기계학습 방법을 적용하였고, 학습 데이터셋에 포함되지 않은 피싱 URL을 분류함으로써 해당 분야에의 기계학습 방법의 타당성을 검증하였다 [8].

A. Le et al. 에서는 보다 복잡한 비선형적인 맵핑을 수행할 수 있다고 알려진 Support Vector Machine (SVM) 을 사용하여 기계학습 기반의 피싱 URL 분류 시스템의 성능을 보완하였고 [9], R. Verma 에서는 URL의 어휘적 특징에 대해 계층적인 요소를 효과적으로 모델링할 수 있는 Random Forest 알고리즘을 도입하는 것으로 분류 정확도를 크게 향상시켰다 [10].

A. Bahnsen et al. 그리고 A. Zhao 에서는 딥러닝 알고리즘 중 통계적인 의미에 기반하여 단어 벡터를 임베딩할 수 있는 Word-to-vector 모형을 사용하여 피싱 URL 으로부터 의미적인 특징을 추출하였고, 게이트 연산을 포함하는 시계열 모델링에 특화된 딥러닝 알고리즘인 LSTM 과 GRU를 사용하여 분류함으로써 해당

분야에의 모델링 방법에서의 개선을 도모하였다 [11],[12].

A. Anand et al. 에서는 데이터의 잠재공간을 효과적으로 모델링할 수 있는 딥러닝 알고리즘인 생성적 적대 신경망에 기반하여 피싱 URL 의 구문적 특징을 생성하고 모델링하는 시도를 통해 해당 분야에서 알려진 데이터 불균형 문제를 해결하고자 하였다 [13]. W. Yang 에서는 Word-to-vector 모형으로부터 추출한 의미적 특징을 효과적으로 모델링하기 위한 컨볼루션-순환신경망을 제안하였다 [14].

그러나 전술한 시도에서는 URL의 구문적/어휘적 특징을 추출하는 과정과 수준별 특징을 융합하는 과정에서 발생하는 성능저하 문제가 있다. 본 논문에서는 기존 시도와 달리 피싱 URL 의 수준별 모델을 융합하는 방법을 학습하는 퓨전 신경망을 제안함으로써 데이터로부터 효과적인 앙상블 규칙을 찾아내기 때문에 피싱 URL 의 분류에 적합하다.

Ⅲ. 제안하는 방법

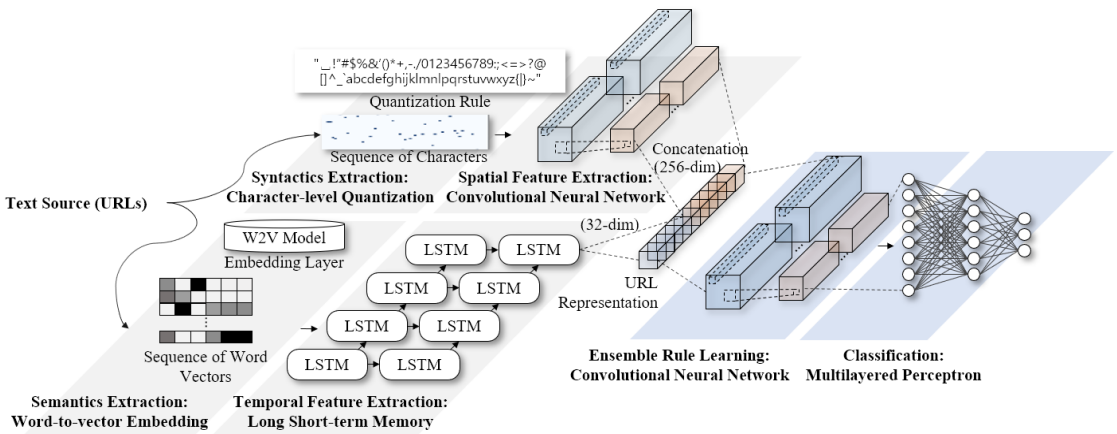
그림 2에서 피싱 URL 분류 분야에서 수준별 특징을 추출하는 딥러닝 모형과 그 융합을 위한 컨볼루션 기반 퓨전 신경망을 도식화하였다. 직관적으로 문자 수준과 단어 수준의 특징을 각각 모델링하고 융합 방법을 데이터로부터 학습하는 시도로 이해할 수 있다.

3.1. 딥러닝 기반 피싱 URL 수준별 특징 모델링

피싱 URL의 구문적인 특징과 의미적인 특징을 각각 모델링하기 위해서 두 가지의 딥러닝 알고리즘과 각각의 전처리과정을 수행하였다. 첫번째, 피싱 URL 에서 자주 관측되는 특수문자를 포함한 임의의 문자의 나열의 구문적 특징을 모델링하기 위해서 문자 각각에 정수를 부여하여 구성한 저수준 신호를 컨볼루션 신경망으로 모델링한다. 두번째, URL 내부를 구성하는 도메인과 서브도메인으로부터 의미적인 특징을 모델링하기 위해서 단어 각각을 Word-to-vector 모형으로 임베딩하여 구성한 단어의 시퀀스를 LSTM 순환 신경망으로 모델링한다.

컨볼루션 신경망은 데이터로부터 필터를 학습하고 입력된 벡터의 패턴으로부터 태스크에 유효한 특징을 추출하는 딥러닝 알고리즘으로, 텍스트의 분류 분야에서 그 강건성과 성능이 입증되었다 [15]. 각 문자는 UTF-8 인코딩 아래에서 고유한 유니코드 값으로 대체하였고, 수집한 데이터셋의 URL 문자 길이의 평균을 고려하여 최대 100자의 정수 시퀀스를 추출하였다. 사용된 문자의 종류는 총 139가지로 문자 수준 컨볼루션 신경망에는 n개의 입력벡터에 대한 $(n \times 100 \times 139)$ 차원의 벡터가 입력된다.

수식 (1)의 컨볼루션 연산 $\phi_c^l(\cdot)$은 파라미터화한 필터를 입력벡터에 대해 적용하고 태스크에 유효한 특징을 추출하는 연산으로 l번째 층의 i행 j열 노드에 대해 $m \times m$크기의 필터를 적용한다:



〔그림 2〕 피싱 URL의 구문적 모델링을 위한 문자수준 컨볼루션 신경망과 의미적 모델링을 위한 단어수준 순환 신경망의 앙상블 규칙을 학습하기 위한 컨볼루션 기반 퓨전 신경망.

$$\phi^{l_c}(x_{ij}) = \sum_{a=0}^{m-1}\sum_{b=0}^{m-1} w_{ab} x_{(i+a)(j+b)} \quad (1)$$

l번째 풀링층에서 수행되는 풀링 연산$\phi_p^l(\cdot)$은 $N\times N$크기의 입력벡터 중 $k\times k$ 영역을 대상으로 풀링 거리 τ에 대해 수식 (2)의 대표값 추출 연산을 수행하고 해당 영역으로부터 최대값을 출력한다:

$$\phi_p^l(x_{ij}) = \max_{\gamma\in R} x_{ij\times\tau} \quad (2)$$

컨볼루션 연산의 학습은 특징 간의 공간적 관계를 보존하며 상관관계를 추출하는 필터 w의 가중치를 찾는 과정이고, 풀링 연산은 왜곡되거나 강조된 특징들로부터 피싱 URL 분류에 유의미한 특징을 추출하는 방식으로 계산복잡도를 감소시키는 차원 감소과정을 의미한다. 제안하는 문자수준 컨볼루션 신경망의 컨볼루션-풀링 모듈은 총 3개 층으로 각자 64개의 필터수와 (2×2)필터 크기를 사용한다.

한편 URL을 구성하는 대표적인 특징중에는 도메인과 하위 도메인의 구성과 같은 단어의 시퀀스로부터 도출할 수 있는 의미적인 특징이 있다. 문자의 시퀀스를 추가적으로 모델링하는 딥러닝 알고리즘을 병렬적으로 사용함으로써 피싱 URL 분류 태스크의 분류 정확도를 보완할 수 있다.

우리는 Word-to-vector 모형을 사용한 단어 임베딩 [16]과 시계열 모델링을 위한 Long Short-term Memory (LSTM) 딥러닝 알고리즘으로 피싱 URL의 의미적인 특징을 모델링한다. 피싱 URL 이 일반적으로 매우 많은 하위 도메인을 포함하고 있는 배경으로부터 우리는 20개의 하위 도메인에서 등장하는 단어들을 추출하였다. 각 단어는 Word-to-vector 모형에 의해 32차원의 벡터로 대체되었고, 피싱단어 수준 순환신경망에는 n개의 입력벡터에 대한 $(n\times20\times32)$차원의 벡터가 입력된다.

LSTM 네트워크는 내부에 비선형적인 세가지의 게이트를 도입한 순환신경망의 일종으로 입력되는 피싱 URL을 구성하는 단어의 시퀀스로부터 선택적인 기억용량에 기반한 수식 (3)의 시계열 모델링 연산 $\phi_L^l(\cdot)$을 수행한다:

$$\phi_L(x_{ij}) = o_t \odot \tanh(c_t) \quad (3)$$

사용되는 입력 게이트 (i), 망각 게이트 (f) 및 출력 게이트 (o) 그리고 LSTM 셀 상태 c 는 ω길이의 입력 단어 시퀀스 $x=(x(t),...,x(t-\omega))$에 대해 수식 (4)와 같이 정의한다. $b,\sigma,\odot$은 각각 신경망에 더해지는 편향과 신경망의 시그모이드 활성화함수 그리고 하다마르곱을 표현한다:

$$\begin{aligned} i_t &= \sigma(W_{ix}x(t) + W_{im}h_{t-1} + b_i) \\ f_t &= \sigma(W_{xf}x(t) + W_{hf}h_{t-1} + b_f) \\ o_t &= \sigma(W_{xo}x(t) + W_{ho}h_{t-1} + b_o) \\ g_t &= \tanh(W_{xh}x(t) + W_{hc}h_{t-1} + b_o) \\ c_t &= f_t \odot c_{t-1} + i_t \odot g_t \end{aligned} \quad (4)$$

3.2. 퓨전 신경망 기반 앙상블 규칙의 학습

제안하는 딥러닝 기반 수준별 피싱 URL의 분류 모형에서는 기존의 수준별 딥러닝 앙상블에서 각 모형이 실제 분류를 수행한 후 분류결과를 평균 또는 투표하던 것과는 달리, n개의 입력에 대한 $(n\times256)$ 그리고 $(n\times32)$ 크기의 분류가 수행되지 않은 중간층 벡터를 활용한다. 문자수준 컨볼루션 신경망과 단어수준 순환신경망에서 출력된 피싱 URL의 구문수준 그리고 단어수준의 분산 표현은 접합되어 $(n\times288)$ 크기의 벡터를 구성하고, 제안하는 퓨전 신경망은 입력된 벡터를 일반 또는 피싱 URL 에 맵핑하는 오차를 최소화하도록 학습한다.

수준별 특징을 체계적으로 융합할 수 있도록 고안된 퓨전 신경망은 컨볼루션 층을 포함하고 있어 288차원의 입력벡터로부터 유효한 특징을 선택한다. 실제 분류는 간단한 신경망으로 수식 (5) 에서 수행한다:

$$p(\hat{y}|x_i) = \arg max \frac{\exp(\phi^{l-1}(x_i)w^l + b^l)}{\sum\exp(\phi^{l-1}(x_i)w + b^l)} \quad (5)$$

신경망의 활성화함수 Softmax 는 출력 벡터가 [0,1] 크기의 확률로써 인코딩되도록 하면서 손실함수의 최적화 과정에서 수행되는 미분과정의 편의를 돕는다. 퓨전 신경망을 포함하는 문자수준, 의미수준 신경망은 입력에서 출력까지 소요되는 모든 함수가 미분 가능하며 데이터로부터 End-to-end 방식으로 학습한다.

〔표 2〕 퓨전 신경망 기반 피싱 URL 분류 혼동행렬.

		Prediction		
		Benign	Phishing	Recall
Actual	Benign	9035	74	0.9919
	Phishing	115	2776	0.9602
	Precision	0.9874	0.9740	

전체 신경망은 수식 (6)의 크로스 엔트로피 손실 함수 L_{CE}에 대해 경사 하강법에 기반한 역전파 알고리즘으로 전체 가중치를 튜닝한다:

$$L_{CE} = -\sum_{i} y_i \log(\hat{y_i}) \tag{6}$$

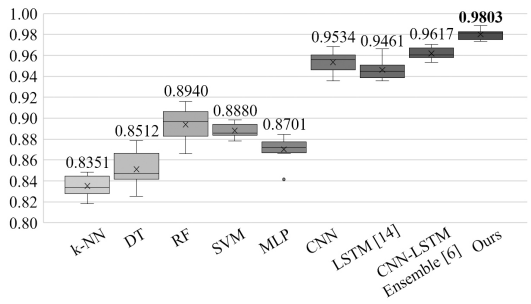
요약하여 제안하는 퓨전 신경망에 기반한 수준별 특징의 융합은 규칙에 기반한 수준별 피싱 URL 특징의 활용 대신, 비선형적 맵핑기능에서 잘 알려진 컨볼루션 신경망에 기반하여 수준별 특징을 분류에 활용함으로써 앙상블 규칙을 학습한다.

Ⅳ. 실험 결과

4.1. 피싱 URL 분류 데이터셋 수집

제안하는 수준별 분류기의 융합을 위한 퓨전 신경망의이 학습한 비선형적인 앙상블 규칙을 검증하기 위해서 기본적인 이진 분류 실험을 설계하였다. 우리는 피싱 URL의 블랙리스트화를 위한 데이터베이스가 공유되는 Phishtank로부터 15,000개의 피싱 URL을 수집하였고 [17], 정상적인 URL의 카테고리화를 위한 데이터베이스가 공유되는 DMOZ-ODP로부터 45,000개의 정상 URL을 수집하였다 [18]. 피싱 URL이 상대적으로 수가 적은 상황을 반영하기 위해 의도적으로 데이터 수를 불균형하게 조정하였다. 웹 크로울러의 규칙에 따라 수집된 URL 데이터셋의 특징의 차이가 발생할 수 있다는 점에서 추후 URL 수집의 범위를 확대하는것이 필요하다.

피싱 URL의 평균 길이가 75.74자임에 비해 정상 URL 평균 길이는 35.83자인 측면에서 피싱 URL 내부의 구문적 특징 모델링이 특히 필요하다. 한편 피싱 URL의 대부분이 삭제된 링크임을 감안할 때 기계학습 기반의 일반화성능이 향후의 피싱 URL 분류에 유망하다.



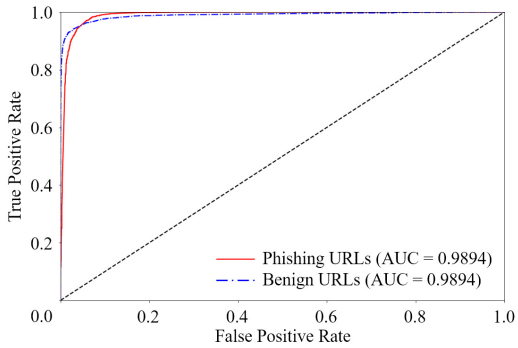
〔그림 3〕 기존 기계학습 방법과의 분류 정확도 비교를 위한 10겹 교차검증.

4.2. 피싱 URL 분류 정확도

제안하는 End-to-end 방식의 피싱 URL 분류 성능을 검증하기 위해서 그림 3에서 기타 기계학습 기반의 피싱 URL 분류 방법들을 포함하여 10겹 교차검증을 수행하였다. 기존의 기계학습 방법들 중 Random Forest 알고리즘이 0.8940의 성능을 나타냄에 비해 컨볼루션 신경망과 순환 신경망이 각각 0.9534와 0.9461의 평균 정확도 향상을 보인다.

Kim 에서 제안한 컨볼루션 신경망과 순환 신경망의 앙상블은 0.9641으로 수준별 특징을 개별로 모델링하고 융합하는 것이 피싱 URL분류 문제에서 유의미함을 보였다. 컨볼루션 신경망을 사용하여 수준별 특징을 융합하고 앙상블 규칙자체를 학습하도록 고안된 퓨전 신경망 기반 피싱 URL 분류 정확도는 0.9803으로 기존 최고성능을 상회하는 결과를 도출하였다.

그림 4에서는 False Negative 오류를 최소화하는것이 필수적인 피싱 URL 분류 분야의 특성을 반영하기



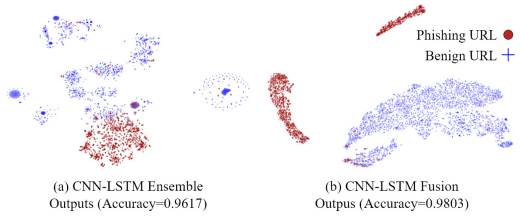
〔그림 4〕 퓨전 신경망 기반 분류결과의 ROC 커브 및 AUC.

위해 ROC커브 및 AUC를 표현하였고, 표2에서 최고 정확도를 기록한 모형에 대한 분류결과를 기재하였다. 실제 피싱 URL을 정상 URL으로 분류하는 오분류 케이스와 피싱 URL의 Recall 이 0.9602인데서 피싱 URL에 빈번한 임의의 문자 나열로부터 발생하는 변산 요소를 추가적으로 모델링하는 것이 필요하다.

4.3. 퓨전 신경망 성능

표 3에서 수준별 딥러닝 모형의 개별 성능을 검증하기 위해 실제 분류사례를 들어 각 모형의 강점을 보였다. 상위 두가지 케이스는 문자 수준 컨볼루션 신경망의 경우 임의의 문자 나열에 대한 강건성을 보인다. 피싱 URL의 구성이 다수의 하위 도메인과 특수문자로 구성되어 있음을 고려할 때 문자수준 컨볼루션 신경망에서는 0.9874의 확률로 피싱 URL으로 분류하였다. 한편 두번째 분류 케이스에서 특정 단어로 인해 단어수준 순환신경망이 정상 URL을 0.8441의 확률로 오분류하는 것에 비해 문자수준 컨볼루션 신경망의 강건성을 확인하였다.

단어수준의 순환 신경망은 문자수준의 컨볼루션 신경망에서 사용하지 않는 하위 도메인을 단어의 형태로 반영함으로써 전체 시스템을 보완하는 역할을 수행한다. 표 3의 세번째와 네번째 케이스에서는 'security', 'bitcoin' 등의 특정 단어들이 피싱 URL에 자주 등장하는 배경지식을 반영하여 순환 신경망에 의해 분류되었다. 이때 컨볼루션 신경망은 하위 도메인의 수에 따라 정상 URL을 비정상으로 오분류하는 것을 확인하였고,



〔그림 5〕 기존 수준별 딥러닝 모형 앙상블 출력값과 퓨전 신경망 출력값의 t-SNE 군집화 비교.

네가지 케이스로부터 문자수준 그리고 단어수준 딥러닝 신경망의 상호 보완성을 정성적으로 확인하였다.

그림 5에서는 기타 기계학습 기반 피싱 URL 분류 방법 중 우수한 성능을 보인 수준별 딥러닝 앙상블에서 출력되는 활성화함수값과 제안하는 퓨전 신경망에서 출력되는 활성화함수값을 동일한 차원축소 알고리즘을 통하여 2차원에 맵핑하였다. 앙상블 규칙까지 함께 학습하는 제안하는 퓨전 신경망 내부에서는 비선형적으로 수준별 피싱 URL 특징이 융합되기 때문에 기존 딥러닝 알고리즘에 비해 2% 이상 성능이 향상되었다.

V. 결 론

본 논문에서는 피싱 URL의 분류 분야에 문자수준 컨볼루션 신경망과 의미수준 순환 신경망을 소개하였고, 각 모형으로부터 추출된 URL의 구문적인 특징과 의미적인 특징을 체계적으로 융합하기 위한 퓨전 신경망을 제안하였다. End-to-end 방식으로 학습되는 퓨전 신경망은 기존에 사용하던 출력값 평균 방식의 앙상블

〔표 3〕 문자수준 컨볼루션 신경망과 단어수준 순환 신경망의 케이스 분석에 기반한 상호 보완성의 정성적 평가.

	Class	URL	CNN Score	LSTM Score
CNN Advantages	Phishing	https://1drv.ms/xs/s!AhtvzT3KrwqMZzLMKnTc8clHnRA?wdFormId=%7BA0F7982D%2D71A4%2D4DE0%2DB4C4%2DC16A0F044EA0%7D	0.9874	0.7385
	Benign	http://market.security***.net	0.0031	0.8441
LSTM Advantages	Phishing	http://bitcoin24-wallet.site	0.0722	0.9837
	Benign	http://www.knightfeatu***.com/kfweb/content/features/kffeatures/puzzlesandcrosswords/kf/sudoku/sudoku_classic/sudoku_classic.html	0.8384	0.0073
Misclassified	Benign	http://archives.seattletimes.nwsou***.com/cgi-bin/texis.cgi/web/vortex/display?slug=will&date=199903	0.8815	0.8764
	Phishing	http://tesla-present.site/ethereum/	0.0584	0.0354

규칙을 파라미터화하여 데이터로부터 학습하는 효과를 보였고, 딥러닝 알고리즘 기반 URL 분류 방법을 포함하여 기계학습 방법 중 최고의 분류정확도 (0.9804) 를 보였다.

학습 및 테스트 데이터셋으로 45,000건의 정상 URL과 15,000건의 피싱 URL을 수집하였고, 정량적 검증으로 10겹 교차검증과 ROC커브, 정성적 검증으로 오분류 케이스와 딥러닝 내부 파라미터를 시각화하여 분석하였다.

향후 연구로써 피싱 URL 분류 분야의 도메인 지식을 보다 반영하는 것이 필요하다. 특히 대부분의 링크가 삭제되었으며 새롭고 복잡한 링크가 등장하는 피싱 URL의 Zero-day 공격 특징을 반영하기 위해서는 Zero-shot, One-shot Learning과 같은 최신의 딥러닝 알고리즘을 도입하는 것이 면밀히 검토되어야 한다. 또한 본 논문에서는 자연어 처리 분야에서 대표적으로 구분되는 구문론적, 의미론적인 특징을 따로 모델링하고 퓨전 신경망을 통하여 융합하였으나, 피싱 URL 분류 분야만의 고유한 수준별 특징을 찾아 모델링하는 것이 추가적으로 고려되어야만 한다.

참 고 문 헌

[1] H. J. Kim, “Image-based malware classification using convolutional neural network,” *Advances in Computer Science and Ubiquitous Compuging*, pp. 1352-1357, 2017.

[2] V. Suganya, “A review on phishing attacks and various anti-phishing techniques,” *Int. Journal of Computer Applications*, vol. 139, pp. 20-23, 2016.

[3] K. L. Chiew, K. S. C. Yong and C. Tan, “A survey of phishing attacks: their types, vectors and technical approaches,” *Expert Systems with Applications*, vol. 106, pp. 1-20, 2018.

[4] I. Qabajeh, F. Thabtah and F. Chiclana, “A recent review of conventional vs. automated cybersecurity anti-phishing techniques,” *Computer Science Review*, vol. 29, pp. 44-55, 2018.

[5] J. Y. Kim, S. J. Bu and S. B. Cho, “Zero-day malware detection using transferred generative adversarial networks based on deep autoencoders,” *Information Sciences*, vol. 460, pp. 83-102, 2018.

[6] H. J. Kim, “Malware classification using convolutional and recurrent neural network,” In Summer Annual Conf. of the Institute of Electronics and Information Engineering, pp. 1329-1331, 2017.

[7] P. Prakash, M. Kumar, R. R. Kompella and M. Gupta, “Phishnet: Predictive blacklisting to detect phishing attacks,” In Proc. of IEEE Int. Conf. on Computer Communications, pp. 1-5, 2010.

[8] J. Ma, L. K. Saul, S. Savage and G. M. Voelker, “Beyond blacklists: learning to detect malicious web sites from suspicious URLs,” In Proc. of the 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, pp. 1245-1254, 2009.

[9] A. Le, A. Markopoulou and M. Faloutsos, “Phishdef: URL names say it all,” In Proc. of IEEE Int. Conf. on Computer Communications, pp. 191-195, 2011.

[10] R. Verma and K. Dyer, “On the character of phishing URLs: Accurate and robust statistical learning classifiers,” In Proc. of the 5th ACM Conf. on Data and Application Security and Privacy, pp. 111-122, 2015.

[11] A. C. Bahnsen, E. C. Bohorquez, S. Villegas, J. Vargas and F. A. Gonzalez, “Classifying phishing URLs using recurrent neural networks,” In 2017 APWG Symposium on Electronic Crime Research, pp. 1-8, 2017.

[12] J. Zhao, N. Wang, Q. Ma and Z. Cheng, “Classifying malicious URLs using gated recurrent neural networks,” In Int. Conf. on Innovative Mobile and Internet Services in Ubiquitous Computing, pp. 385-394, 2018.

[13] A. Anand, K. Gorde, J. R. A. Moniz, N. Park, T. Chakraborty and B. T. Chu, “Phishing URL detection with oversampling based on text generative adversarial networks,” In 2018 IEEE

Int. Conf. on Big Data, pp. 1168-1177, 2018.
[14] W. Yang, W. Zuo and B. Cui, “Detecting malicious URLs via a keyword-based convolutional gated recurrent unit neural network,” IEEE Access, 2019.
[15] X. Zhang, J. Zhao and Y. LuCun, “Character-level convolutional networks for text classification,” In Advances in Neural Information Processing Systems, pp. 649-657, 2015.
[16] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado and J. Dean, “Distributed representations of words and phrases and their compositionality,” In Advances in Neural Information Processing Systems, pp. 3111-3119, 2013.
[17] OpenDNS, “Phishtank,” http://www.phishtank.com.
[18] NetScape, “Dmoz open directory project,” http://www.dmoz.org.

〈저자소개〉

부 석 준 (Seok-Jun Bu)
2016년 2월 : 한양대학교 컴퓨터공학과 학사
2016년 3월~현재 : 연세대학교 컴퓨터과학과 석박통합과정
<관심분야> 신경망, 진화연산, 베이즈통계

김 혜 정 (Hae-Jung Kim)
2005년 3월~2016년 2월 : 계명대학교 조교수
2016년 3월~현재 : 경일대학교 사이버보안학과 부교수
<관심분야> 사이버보안, 인공지능응용, 딥러닝