Check for updates

Journal of KIIT. Vol. 21, No. 7, pp. 63-69, Jul. 31, 2023. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2023.21.7.63

멀티노미얼나이브베이즈 기법의 정교화를 통한 악성 메일 필터링 시스템 구현

지경엽*, 권영미**

Implementation of Malicious Mail Filtering System through Refinement of MultinomialNB Technique

Keungyeup Ji*, Youngmi Kwon**

요 약

악성 메일의 비중이 점차 심각하게 증가하는 상황에서 악성 메일 필터링의 정확도 향상을 위해서 메시지 규칙기반 방법 대신 세가지 유형의 기계학습 필터링 방법을 본 논문에 적용하였으며, Naive Bayes 알고리즘인 멀티노미얼나이브베이즈 기법이 다른 두 가지 기계학습 기법보다 악성 메일 예측정확도가 더 우수하고 처리시간도 더 적게 소요된다는 것을 입증하는 것이 주요 목표이다. 이를 위해서 1,454,489건의 데이터를 활용하여 실험한 결과에 의하면 스팸 예측오류율이 MultinomialNB 적용 결과는 8%이고, SVM 결과는 42%, LR 결과는 20%로 나왔으며, 실행시간 결과는 MultinomialNB 적용 시 1,489초, SVM 결과는 10,301초, LR 결과는 1,963초가 소요되었다. 결론적으로 효율적인 악성 메일 필터링 구현을 위해서는 MultinomualNB 알고리즘 기반하에 라플라스 스무딩과 적절한 알파 파라미터값을 적용할 것을 권장한다.

Abstract

In a situation that malicious e-mails are seriously increased, in order to increase accuracy of malicious mail machine learning methods of three kinds instead of message rule-based methods were applied to this paper. The main goal is to prove that the multinomial Naive Bayes technique of a Naive Bayes algorithm has better prediction accuracy for malicious emails and takes less processing time than other two kinds of machine learning techniques. To prove this, according to the experimental result consisting of 1,454,489 the spam prediction error rate was 8% for the MultinomialNB algorithm 42% for the SVM algorithm and 20% for the LR algorithm. In conclusion, I suggest that implementing a malicious mail filtering system by applying the Laplace smoothing technology and good alpha parameter value based on the MultinomualNB algorithm is an effective method.

Key words
naive bayes, SVM, LR, multinomialNB, Laplace smoothing

* 충남대학교 전자정보통신공학과 박사
- ORCID: https://orcid.org/0000-0002-6610-1053
** 충남대학교 전파정보통신공학과 교수(교신저자)
- ORCID: https://orcid.org/0000-0003-0318-0660

+ Received: May 22, 2023, Revised: Jun. 05, 2023, Accepted: Jun. 08, 2023
+ Corresponding Author: Youngmi Kwon
Dept. of Radio and Info. Communications Eng., Chungnam National University
Tel.: +82-42-821-6890, Email: ymkwon@cnu.ac.kr

Ⅰ. 서 론

비즈니스 거래나 개인 간의 교신으로 메일의 중요성은 증가하고 있으며 법적 효력도 있는 상황에서 스팸메일은 점차 증가하고 있으며, 통계적으로 전체 메일의 45% 이상을 차지하고 있다[1]. 스팸메일 중 광고성 스팸메일은 36% 이상이고 그중에서 신용사가 관련 스팸메일은 2.5%이다. 특히 개인에게 심각한 재산 피해와 개인정보 유출을 하는 피싱메일은 신용사기 스팸메일 중에서 73% 이상을 차지하고 있다[1]. 스팸 필터링을 하는 방법으로는 메시지 규칙 기반방식으로 데이터베이스에 블랙 리스트나 화이트 리스트 정보를 미리 등록하여 차단 또는 허용하는 방식이 있다. 이의 문제점은 데이터베이스에 관련 정보를 미리 저장해야 효과가 있는 것이다. 이의 문제를 해결하고자 스팸으로 다수 신고된 단어를 해시값으로 데이터베이스에 저장하여 차단하는 방식도 있으나[2] 기존 방식은 데이터베이스에 스팸 정보를 미리 등록해야 차단 효과가 있는 단점이 존재한다. 이를 해결하고 위해서 스팸메일을 보다 효율적으로 필터링하려는 방법으로 본 논문은 기계학습방법 중에서 MultinomialNB를 활용하여 필터링의 정확도를 높였으며 다른 기계학습 방법인 SVM(Support Vector Machine)과 LR(Logistic Regression) 방법으로도 스팸 필터링을 하여 비교분석을 하였다. 본 논문의 구성은 II 장에서 관련 연구를 하였으며 III 장에서는 나이브베이즈 기계학습 방법을 활용하여 스팸 필터링 시 예측오류율을 낮추기 위한 방법을 제시하였고, IV 장에서는 나이브베이즈 기법에 속한 MultinomialNB와 SVM, LR 기계학습 방법 간의 스팸 필터링 시 예측오류율 및 정확도 결과를 비교분석 하였다. V 장은 결론으로 구성하였다.

Ⅱ. 관련 연구

2.1 기계학습의 개요

기계학습은 패턴과 추론을 기반으로 학습을 통해서 판단력을 향상하는 알고리즘이며 향상된 판단력으로 예측오류율을 감소시킨다[3]. 기계학습은 지도기계학습(supervised machine learning), 자율기계학습(unsupervised machine learning) 및 강화 기계학습(reinforcement machine learning)으로 구분된다. 지도학습 개념은 기준이 되는 목표값과 판단을 위한 데이터가 존재하며, 분류(classification)와 회귀(regression)가 속한다. 자율학습은 목표값은 존재하지 않고 판단을 위한 데이터는 존재하는 경우가 해당하며, 강화학습은 목표값과 데이터가 존재하지 않고 보상을 기반으로 학습하는 것이다[3].

2.2 나이브베이즈 알고리즘

나이브베이즈 알고리즘은 베이지안 이론에 근거를 둔다. 베이지안 이론은 사건들이 상호 의존적이고 독립적인 관계라는 가정하에 사후 확률을 구한다[3]. 베이지안 이론 개념은 사건들이 상호 의존적이고 각 사건은 다른 사건에 영향을 받지 않는 독립적인 관계라는 가정하에 사후 확률을 의미한다. 베이지안 이론 식은 식 (1)에 표현되었다.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

- P(A): 사전확률. 사건 A의 확률
- P(B|A): 가능도 확률(Likelihood probability). 원인 A가 발생되었다는 가정하에서 결과 B가 발생할 확률
- P(A|B): 사후 확률. 결과 B가 발생하였다는 가정하에서 원인 A가 발생할 확률

식 (1)에 의해서 나이브베이즈 알고리즘은 복잡하게 여러 개의 특성으로 얽혀있는 그룹을 유사한 성격끼리로 재분류할 수 있는 특성을 보인다[4]. 나이브베이즈 알고리즘의 응용 분야로는 이메일 스팸분류, 네트워크 침입 탐지, 질병 발생 탐지 등에 활용된다.

2.3 라플라스 스무딩(Laplace smoothing)

라플라스 스무딩 기술은 나이브베이즈 알고리즘 기법중의 하나인 멀티노미얼나이브베이즈(Multinomi

alNB)에 적용하는 기술로서 확률 계산이 0이 되는 문제를 해결함으로써 알고리즘 운영을 안정적으로 수행하는 것이 주요 목적이다. 라플라스 스무딩 식은 식 (2)에 기술하였다[3][5].

$$P(w'|positive)=\frac{w'\text{발생수}(y=positive)+\alpha}{N+\alpha * K} \quad (2)$$

- alpha: 스무딩 파라미터
- K: 데이터 차원 개수
- N: positive 상태에서 전체 발생 개수

2.4 SVM 알고리즘

SVM은 통계학습이론을 기반으로 한 기계학습 알고리즘이다[6]. 이 논문은 SVM 기계학습 기법을 Hadoop MapReduce 프레임워크에서 적용하여 스팸 필터링을 수행하는 내용을 기술하였다. SVM은 분류 또는 회귀 문제 모두에 사용할 수 있는 지도 기계학습 알고리즘(Supervised machine learning)이다[7][8]. SVM 기법의 응용은 음성신호 자료 분석, 발전소 이상 탐지 및 시스템 호출 등에 활용된다[9]. 또한, SVM의 데이터 분류의 강점을 활용하여 HTTPS 트래픽이 원문 그대로 암호화된 상태에서 악성코드를 탐지할 방안을 제시하였다[10].

2.5 LR 알고리즘

로직스틱 회귀(Logistic regression) 알고리즘은 로직스틱 함수를 사용하여 0 또는 1의 두 가지 경우의 확률을 추정하는 기법으로서 식 (3)에 기술하였다[11].

$$Ø(z) = \frac{1}{1+e^{-z}} \quad (3)$$

- 그림 1에서 x 축에서 z 가 ∞ 로 향할 때 y축에서 로지스틱 함수 (Ø(z))는 1에 가까워지고, -∞ 로 향할 때 Ø(z)는 0에 가까워짐으로써 결과값은 0 또는 1로 산출된다.

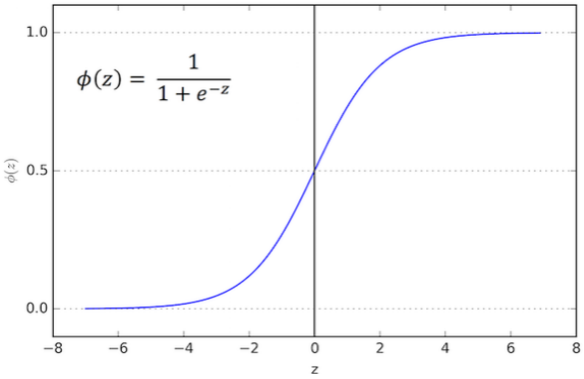

그림 1. Logistic regression 함수
Fig. 1. Logistic regression function

로지스틱 회귀 알고리즘의 주요 응용 분야로는 스팸메일 판단 및 과거의 품질데이터를 학습하여 제품의 불량 여부를 판단하는 데 활용된다.

III. MultinomialNB에 라플라스 스무딩을 적용한 스팸 필터링 방법

[2]에서 나이브베이즈 기법에 속한 MultinomialNB 방법을 활용하여 스팸메일 필터링을 수행하였으며, 본 논문에서는 추가로 SVM 기법과 LR 기법을 활용하여 스팸 필터링을 수행하여 예측오류율과 정확도 산출하였다. 나이브베이즈 기법의 MultinomialNB 방법에서는 오류예측률을 낮추고 예측 정확도 향상을 위해서 예측오류율과 실행시간이 가장 좋은 파라미터 값을 찾기 위해서 12개의 파라미터 값을 활용해서 실행시간과 예측정확도율 및 예측오류율 결과를 산출하였다.

표 1은 multinomialNB의 라플라스 스무딩에 사용되는 알파 파라미터값 유형별로 실행시간을 정리한 것이다[3]. 실험을 위한 데이터 건수는 1,454,489건으로서 알파 파라미터가 0.0001인 경우가 1467.8초로 실행시간이 가장 적게 나왔다.

표 2는 알파 파라미터 유형별로 악성 메일에 대한 예측오류율 및 정확도를 실험한 결과를 기술한 것이다[3]. 알파 파라미터값이 적을수록 예측오류율은 적어지고 정확도율이 상승하였으며, 알파 파라미터값이 증가할수록 악성 메일 예측오류율이 커지면서 정확도율이 감소한다.

표 1. 알파 파라미터 유형별 multinomialNB 실행시간
Table 1. MultinomialNB execution time by alpha parameter type (Unit: sec)

Alpha parameter	Execution time
0	1494.2
0.00001	1468.5
0.00005	1473.2
0.0001	1467.8
0.0005	1492.6
0.001	1490.4
0.005	1500.7
0.01	1474.9
0.05	1482.7
0.1	1471.1
1	1498.6

표 2. 알파 파라미터 유형별 multinomialNB 기법의 악성메일 예측오류율
Table 2. Malicious mail prediction error rate of multinomialNB technique by alpha parameter type

Alpha parameter	Spam prediction error rate	Ham prediction error rate	Total prediction error rate
0.0	3.855%	1.755%	2.412%
0.00001	4.837%	2.202%	3.026%
0.00005	6.831%	3.110%	4.274%
0.0001	7.751%	3.529%	4.850%
0.0005	10.879%	4.953%	6.807%
0.001	12.459%	5.672%	7.796%
0.005	15.884%	7.232%	9.939%
0.01	17.824%	8.115%	11.153%
0.05	21.711%	9.885%	13.585%
0.1	24.351%	11.087%	15.237%
0.5	30.180%	13.741%	18.884%
1.0	31.532%	14.356%	19.730%

표 2에서 측정한 12개의 알파 파라미터 중에서 실행시간과 확률 계산 시 안정성 등을 고려해서 알파 파라미터에 0.0001인 경우를 선택하였다[3]. 표 3은 표 2를 기반으로 예측정확도율을 산출하였다. 표 4는 MultinomialNB 알고리즘과 SVM, LR 알고리즘을 동일한 시스템에서 학습과 예측 시에만 각각의 알고리즘을 실행한 결과이다. 악성 메일 필터링 측정 결과 중 정확도 건수를 측정한 자료로 MultinomialNB 기법이 가장 정확도가 우수한 것으로 결과를 산출하였다. 그림 2는 표 4를 그래프로 표현한 것이다.

표 3. 알파 파라미터 유형별 multinomialNB 기법의 악성메일 정확도
Table 3. Malicious mail prediction accuracy rate of multi nomialNB technique by alpha parameter type

Alpha parameter	Total accuracy ratio
0.0	97.645%
0.00001	97.062%
0.00005	95.901%
0.0001	**95.375%**
0.0005	93.627%
0.001	92.768%
0.005	90.960%
0.01	89.966%
0.05	88.040%
0.1	86.778%
0.5	84.116%
1.0	83.521%

표 4. 기계학습 간 성능평가지표 비교내역
Table 4. Performance evaluation index comparison among machine learning (Unit: count)

Item	MultinomialNB	SVM	LR
Real spam	455,036		
Real Ham	999,453		
TP	**419,767**	263,418	365,693
TN	**1,034,722**	1,191,071	1,088,796
FP	**-35,269**	-191,618	-89,343
FN	**35,269**	191,618	89,343

- TP: 실제 스팸메일을 스팸메일로 예측한 경우 (정답)
- TN: 실제 햄 메일을 햄 메일로 예측한 경우(정답)
- FP: 실제 햄 메일을 스팸메일로 예측(오답)
- FN: 실제 스팸메일을 햄으로 예측(오답)

(Unit: count)

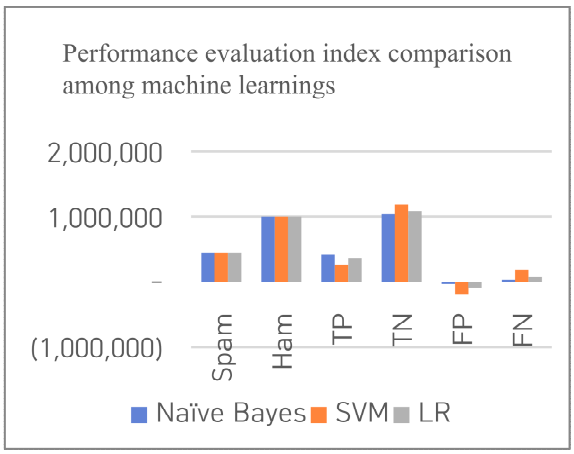

그림 2. 기계학습 간 성능평가지표 비교내역 그래프
Fig. 2. Graph of performance evaluation index comparison between machine learning

표 5는 표 4의 성능평가지표 건수를 참조하여 각 기계학습 방법을 적용하여 악성 메일 예측 시 오류율과 정확도율을 산출한 것이다. 스팸 예측 오류율 결과는 MultinomialNB 알고리즘을 활용했을 때의 결과가 8%이고, SVM 기법의 경우 42%, LR 기법의 경우 20%로 측정되었다. 전체 예측오류율 결과는 MultinomialNB기법 적용 결과가 5%, SVM 기법이 26%, LR 기법이 12%로 측정되어서 Multi nomialNB 기법이 가장 정확한 필터링 결과를 보였다. 결과적으로 전체 정확도 비율은 MultinomialNB기법이 95%의 정확도를 나타냈으며 SVM 기법 적용 결과는 74%의 정확도를 나타냈고, LR 측정기법 적용 결과는 88%의 정확도 결과를 나타냈다.

표 5. 기계학습방법 간 예측오류율 및 정확도 비교
Table 5. Comparison of prediction error rate and accuracy among machine learning methods (Unit: %)

Item	Spam prediction error rate	Ham prediction error rate	Total prediction error rate	Total accuracy rate
Multinomial NB	8	4	5	95
SVM	42	19	26	74
LR	20	9	12	88

그림 3과 그림 4는 표 5의 기계학습의 예측오류율 및 정확도율 수치를 그래프로 표현한 것으로서 Naive Bayes 알고리즘은 MultinomialNB기법을 의미한다.

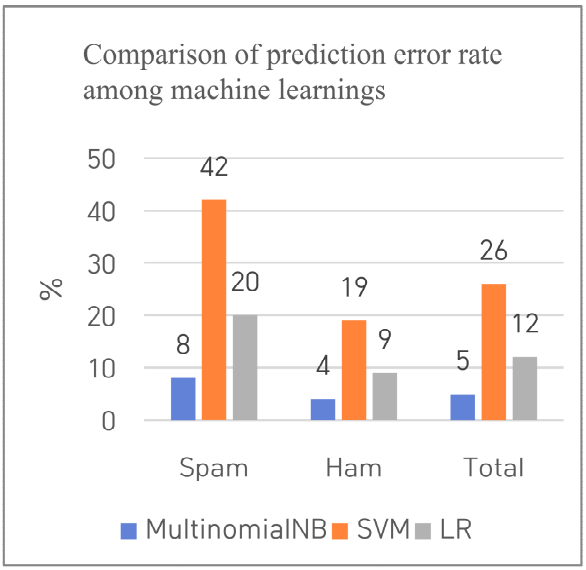

그림 3. 기계학습 간 예측오류율 비교내역
Fig. 3. Comparison of prediction error rates among machine learning

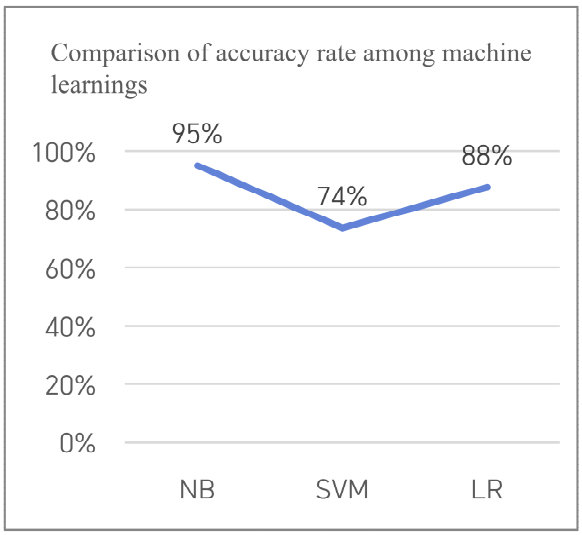

그림 4. 기계학습 간 정확도율 비교내역
Fig. 4. Comparison of accuracy rates among machine learning

기계학습 간 실행시간은 표 6에 기술한 것처럼 MultinomialNB 기법 1,489초가 소요되었고 SVM 알고리즘 적용 결과는 10,301초, LR 알고리즘 적용 결과는 1,963초가 소요되어서 MultinomialNB기법의 실행시간이 가장 적게 산출되었다.

표 6. 기계학습 간 실행시간 비교내역
Table 6. Comparison of execution time among machine learning (Unit: sec)

Technique type	Execution time
Naive bayes	1,489
SVM	10,301
LR	1,963

Ⅳ. 결 론

악성 메일 예측오류율을 감소시키기 위해서 본 논문은 나이브베이즈 기계학습기법 중의 하나인 multinomialNB 기법을 활용하여 학습을 통하여 악성 메일 예측오류율을 감소시키고 정확도를 향상했다. 본 논문에서 적용한 방법은 multinomialNB 알고리즘 기법을 기반으로 라플라스 스무딩 기술을 적용하여 악성 메일에 대한 예측오류율을 감소시켰다. 특히 라플라스 스무딩 기술을 효과적으로 활용하기 위해서 라플라스 스무딩을 제어하는 알파파라미터 값을 실행시간과 예측오류율을 고려해서 가장 적절한 값인 0.0001을 선정하여 실험하였다. 본 논문에서 적용한 알고리즘 기법이 실행시간과 악성 메일에 대한 예측오류율에서 다른 기계학습기법인 SVM이나 LR 알고리즘 기법을 활용하여 산출된 결과와 비교분석을 하여 상대적인 우수성을 입증하였다.

결론적으로 수신된 메일 중에서 악성 메일 여부를 판단하기 위해서 기계학습을 적용 시에는 multinomialNB 알고리즘 기법 기반하에 라플라스 스무딩 기술을 적용하고, 정확도와 성능에 영향을 끼치는 알파 파라미터값에 가장 적합한 값을 선정하여 악성 메일 필터링을 하는 것이 가장 합리적인 방안인 것으로 보인다.

참 고 문 헌

[1] D. Sorkin, "SPAM LAWS", SpamLaws, 2019. https://www.spamlaws.com/spam-stats.html. [accessed: May 30, 2023]

[2] S. Byun and J. Kim, "Spam message filtering system using message digest algorithm", Proceedings of KIIT Conference, pp. 120-123, May, 2014.

[3] K. Ji, "A new malicious mail filtering method using Naïve Bayes technique on efficiently tuned Hadoop framework", Chungnam Nation al University, Daejeon, Korea, pp. 18-20, 74-101, Feb, 2023.

[4] J. Jang, "easy understanding for AI 3-2, Naive Bayes classification algorithm", ZDNET Korea, 2022. https://zdnet.co.kr/view/?no=20220530181623&re=O_20220530181623&p=4. [accessed: May 30, 2023]

[5] V. Jayaswal, "Laplace smoothing in Naive Bayes algorithm", Towards DataScience, 2020. https://towardsdata science.com/laplace-smoothing-in-naive%C3%AFve-bayes-algorithm-9c237a8bdece. [accessed: May 30, 2023]

[6] D. Somvanshi, "A survey on spam filtering methods and Map-Reduce with SVM", International Research Journal of Engineering and Technology(IRJET), Vol. 4, No. 3, pp. 490-494, Mar. 2017.

[7] Shreyak, "Spam Mail Detection Using Support Vector Machine", Becoming Human:Artificial Intelligence Magazine, 2020. https://becominghuman.ai/spam-mail-detection-using-support-vector-machine-cdb 57b0d62a8. [accessed: May 30, 2023]

[8] O. Amayri and N. Bouguila, "A study of spam filtering using support vector machines", Springer Science, pp. 73-108, May 2010. https://doi.org/10.1007/s10462-010-9166-x.

[9] M. Oh, A. Park, Y. Kim, and J Jin, "Research of anomaly detection technique based on machine learning", Korea Institute for Health and Social Affairs, pp. 42-43, Dec. 2019.

[10] D. Jeon and D. Park, "Malware Detection in Encrypted TLS Traffic using Machine Learning Techniques", Journal of KIIT, Vol. 19, No. 10, pp. 125-136, Oct. 2021. https://doi.org/10.14801/jkiit.2021.19.10.125.

[11] Natasha Sharma, "Spam Detection with Logistic Regression", Towards DataScience, 2018. https://towardsdatascience.com/spam-detection-with-logistic-regression-23e3709e522. [accessed: May 30, 2023]

저자소개

지 경 엽 (Keungyeup Ji)

1989년 2월 : 광운대학교 전자계산학과(공학사)
2016년 2월 : 충남대학교 전파정보통신공학과(공학석사)
2023년 2월 : 충남대학교 전파정보통신공학과(공학박사)
관심분야 : Hadoop, Spark, Cloud Computing, Machine Learning, Data Mining

권 영 미 (Youngmi Kwon)

1986년 2월 : 서울대학교 컴퓨터공학과(공학사)
1988년 2월 : 서울대학교 컴퓨터공학과(공학석사)
1996년 8월 : 서울대학교 컴퓨터공학과(공학박사)
1993년 ~ 1995년 : ETRI 연구원
1996년 ~ 2002년 : 목원대학교 컴퓨터공학과 조교수
2006년 ~ 2007년 : Indian Statistical Institute 객원연구원
2002년 ~ 현재 : 충남대학교 전파정보통신공학과 교수
관심분야 : Internet Protocols, WSN, Embedded System, Cloud Computing, Distributed System