가우시안 혼합 모델 군집화를 적용한 안드로이드 악성 앱 탐지 기법

이승민, 안석현, 조성제, 김동재, 황영섭

ASK 2025

2025.05.30

컴퓨터보안 및 OS 연구실

단국대학교 DANKOOK UNIVERSITY

Computer Security & OS LAB

INDEX

01 서론

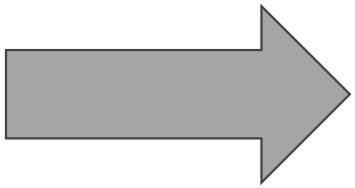
Introduction

❖ 악성 앱 탐지 기법

- 전통적인 방식: Signature-based approach
- 새로운 유형의 악성 앱, 다형성 악성 앱(polymorphic malware)를 탐지하기 어려움

❖ 악성 앱 탐지 회피 기술의 진화

- Code obfuscation
- Packing
- Dynamic code loading



기존 모델의 노후화로 인한 성능 저하

❖ 성능저하 대응 방안

1. 모델 재학습(periodic retraining)
2. 드리프트 탐지 후 모델 업데이트
3. 강건한(robust)한 모델 개발

1.1 전문가의 레이블링(labeling) 작업

- 비용, 시간

2.1 Drift 탐지 정확성 보장 및 업데이트 비용

모델 노후화(Model Aging)

❖ 모델 노후화?

- 과거 데이터로 학습된 모델이 새로운 패턴의 데이터 등장으로 인해 **성능이 저하**되는 현상

[8] L. Tang et al. (2024), "**Demystifying the Evolution of Android Malware Variants**",
IEEE Transaction on Dependable and Secure Computing

- 수동적 진화(Passive Evolution)
  - OS version upgrade, API 정책 변화
  - Android 6.0(Marshmallow) 런타임 권한 요청 → **특징 분포(feature distribution)** 변화

- 능동적 진화(Active Evolution)
  - 악성 앱 탐지 알고리즘 우회 → 새로운 공격 패턴

✓ 이러한 악성 앱의 진화로 **개념 드리프트(concept drift)** 발생

02 개념 드리프트 탐지

Data Collection & Preprocessing

❖ 데이터 수집

- AndroZoo(Luxembourg univ)

Year	Benign	Malware	Total	
2014	5,000	5,000	10,000	Train
2015	5,000	5,000	10,000	
2016	5,000	5,000	10,000	
2017	3,000	3,000	6,000	Test
2018	3,000	3,000	6,000	
2019	3,000	3,000	6,000	
2020	3,000	3,000	6,000	
2021	3,000	3,000	6,000	
2022	3,000	3,000	6,000	
2023	3,000	3,000	6,000	
Total	36,000	36,000	72,000	

❖ 데이터 전처리(Preprocessing)

- Feature: API Calls

[4] Jeong et al.(2024), "**Enhancing Sustainability of an Android Malware Detection Technique using K-means Clustering**", JSAV

[7] 조우상 등(2022), "**연도별 데이터 조합 기계학습을 통한 안드로이드 악성 앱 탐지의 지속 가능성 향상 연구",** 한국차세대컴퓨팅학회논문지

✓ 악성 앱 탐지에 효과적인 **1,848개 API 선정**

✓ 각 앱들의 **API 참조 횟수** 기록

Verify Concept Drift

❖ Used Classifier Models

- RandomForest
- AdaBoost
- KNN

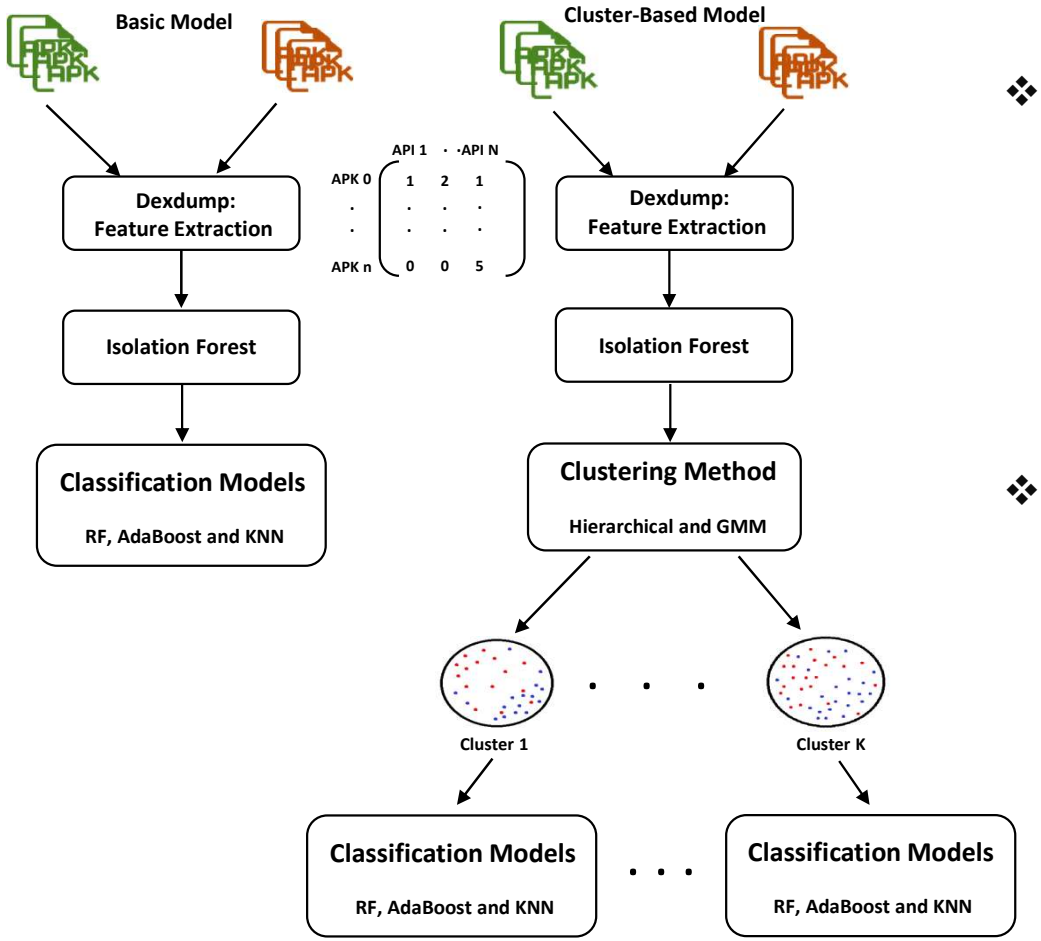
❖ RandomForest **Feature Importance**

Feature Importance	Top 10 API Indexes			
	2014	2017	2019	2021
1	1	1	42	5
2	2	99	157	217
3	3	16	192	309
4	4	2	207	177
5	5	10	240	235
6	6	49	73	170
7	7	67	292	386
8	8	66	115	249
9	9	31	29	60
10	10	126	70	256

03 개념 드리프트에 강인한 악성 앱 탐지 기법

Flowchart

❖ GMM(Gaussian Mixture Model) based Classifier(RF, KNN, AdaBoost)



❖ Experiments

- 1. Basic Model: RF, AdaBoost, KNN
- 2. Hierarchical + (RF, AdaBoost, KNN)
- 3. GMM + (RF, AdaBoost, KNN)

❖ Evaluation Metrics

- F1-score
- AUT(모델의 지속가능성 평가)
  - [22] F. Pendlebury et al(2019). , **"{TESSERACT}: Eliminating experimental bias in malware classification across space and time**", 28th USENIX security symposium
  - 0 ~ 1의 값, 1에 가까울 수록 지속 가능성이 높은 모델

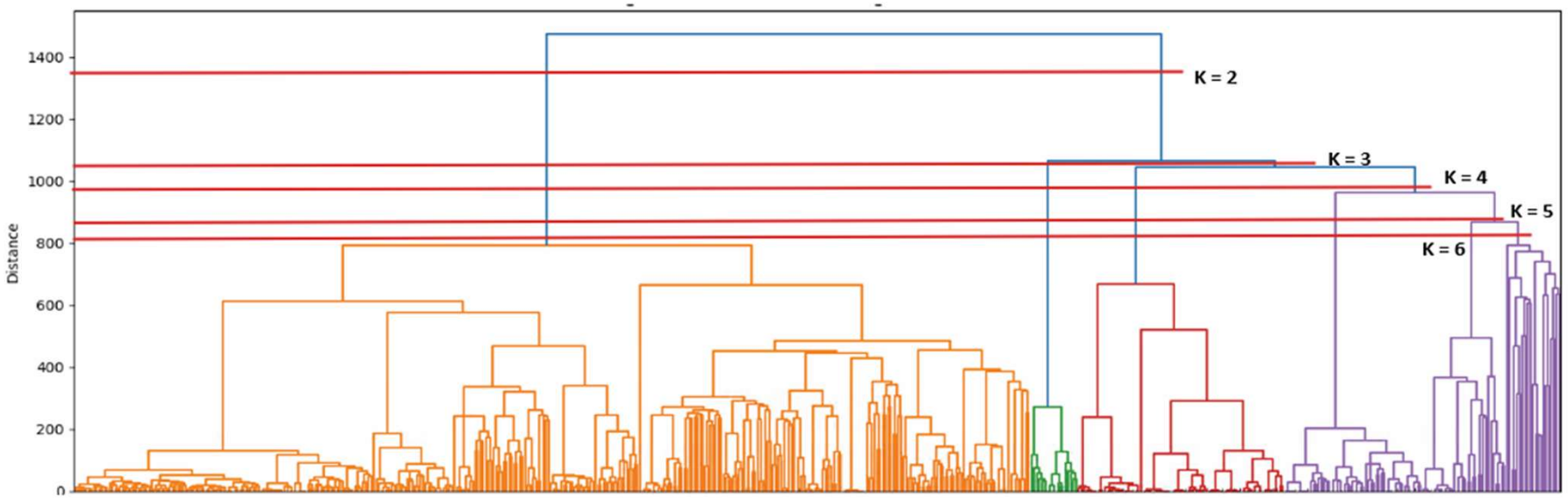
Hierarchical Clustering

❖ 계층적 군집화(Hierarchical Clustering)

- 사전에 군집 수(K)를 지정하지 않음

❖ K 선정 기법

- dendrogram: 데이터 간 유사도 측정
- Silhouette Score(SS): 군집 간 분리도, 응집도 계산
- Within-Cluster Sum of Square(WCSS): 군집 내 샘플의 중심(centroid)부터의 거리 합



K	2	3	4	5	6	7	8	9	10
SS	0.367	0.366	0.347	0.351	0.35	0.05	0.05	0.05	0.05
WCSS(x10^6)	29.58	29.02	28.47	27.98	27.52	27.2	26.88	26.59	26.59

GMM(Gaussian Mixture Model) Clustering

❖ GMM Clustering

- 사전에 K 선정
- AIC(Akaike Information Criterion)
  - **낮을 수록 모델 적합도 상승**

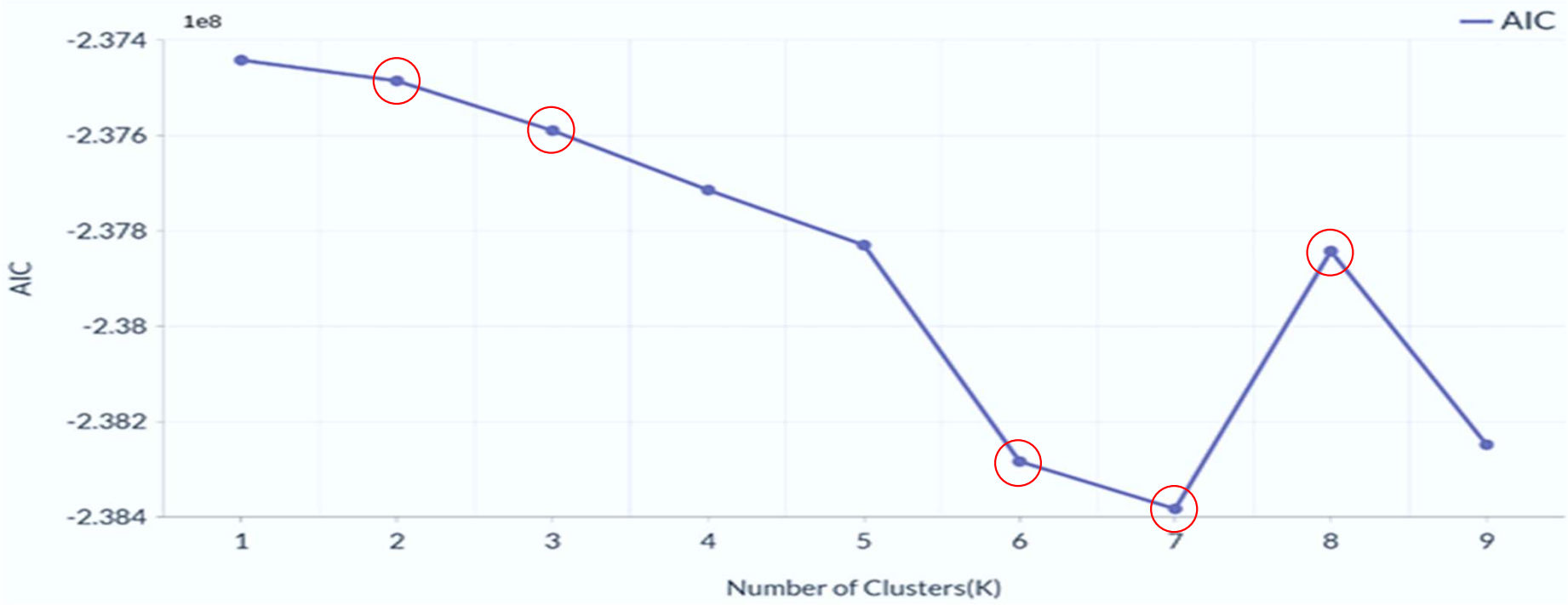
❖ $AIC = 2p - 2ln(L)$

- P: 모델의 파리미터 수
- L: 최대 우도값(likelihood)

❖ K 선정

- local minimum: 7

✓ Over Clustering 고려 **K = 2, 3, 6, 7, 8 선정**



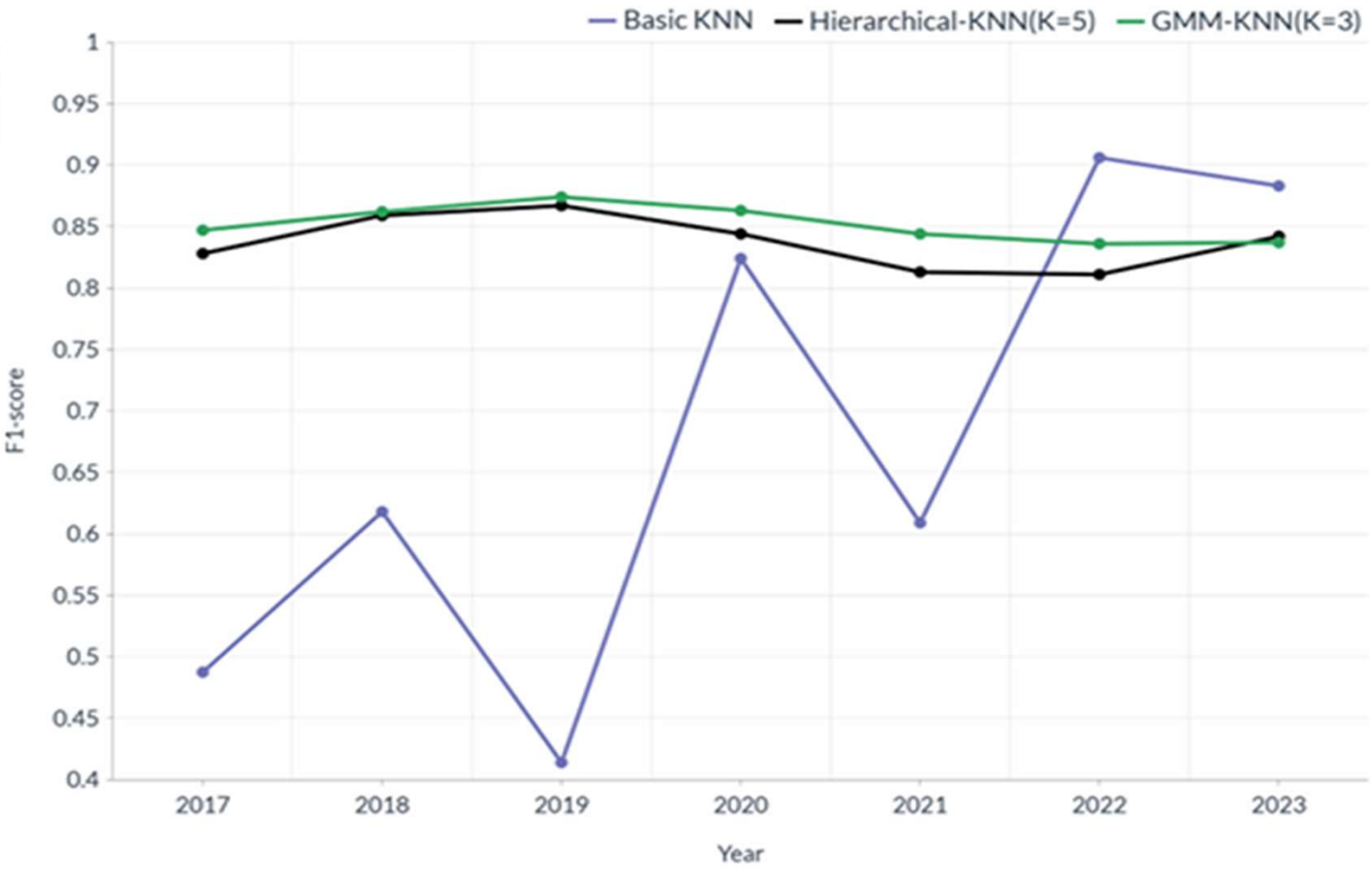
04 실험 및 평가

Experiments Results

❖ Result of Basic Models

Model	RF	
	F1	A
Score	0.765	0.

❖ Result of Hierarchical based Classifier



연도별 모델 성능

erarchical-KNN	
l	AUT
26	0.82
27	0.82
21	0.818
38	0.838
36	0.836

GMM-KNN	
l	AUT
19	0.815
51	0.854
49	0.851
29	0.829
28	0.826

05 결론 및 향후 연구

결론 및 향후 연구

Conclusion

- GMM기반 분류기를 통해 전통적인 모델 대비 F1-score 8.9%p 향상, AUT 10.1%p 향상
- 확률적인 모델링을 기법을 통해 concept drift가 발생하는 2019년 이후 데이터에 대해서도 일관된 성능
- GMM 기반 탐지 기법을 통해 학습되지 않은 유형의 데이터일지라도 탐지할 수 있음

Future Work

- 온라인 메커니즘
- 차원 축소 기법 적용, GMM 계산 효율성 향상
- 개념 드리프트 감지 및 대응 프레임 워크 개발

Acknowledgement

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원 - 대학 ICT연구센터(ITRC)의 지원(IITP-2025-RS-2023-00259967),

산업통상자원부(MOTIE)와 한국에너지기술평가(KETEP) 지원(No.RS-2021-KP002461),

2024년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음(2024-0-00035).

Thank you

Q&A

Appendix #1

❖ Hyperparameter Each Model

Model	Type	Hyperparameter	Value
RF	Classification	n_estimators	100
		max_depth	None
KNN	Classification	n_neighbors	1~5
		weights	uniform
AdaBoost	Classification	n_estimators	50
		learning_rate	1.0
Hierarchical	Clustering	similarity metric	euclidean distance
		linkage	ward
GMM	Clustering	reg_covar	1e-4
		tol	1e-5

❖ Training Time

K	Basic KNN	Hierarchical-KNN	GMM-KNN
Time	10.98 sec	506.99 sec	95.6 sec

❖ IsolationForest Result

Year	Removed	Reamin
2014	48	9952
2015	146	9854
2016	560	9440
Total	754	29426

Concept Drift Related Works

Characteristics	Jeong[4]	Xu[11]	Cai[12]	Mariconti[13]	Zhang[14]
Dataset	Androzoo[15]	Androzoo[15]	Androzoo[15], VirusShare, Google Play	Google Play, Drebin VirusShare	VirusShare, VirusTotal, AMD, Google Play
Data Period	2014 ~ 2016, 2019 ~ 2021	2011 ~ 2016	2010 ~ 2017	2010.10 ~ 2016.05	2012 ~ 2018
Model	K-means, RF	PA, OGD, AROW, RDA, Ada-FOBOS	RF	RF	Mariconti[13], Xu[11], DREBIN[23], DREBIN-DL[24]
Analysis Method	Static	Static	Dynamic	Static	Static
Feature Info	API Call	API Call	SAD Profile	API Call Sequence	API Call