KOMPYUTER VIRUSLARINI KLASSIFIKATSIYA QILISHDA MATEMATIK KLASTERLASH ALGORITMLARINING ROLI

Axmedova Feruza G‘ayrat qizi
axmedovaferuza2004@gmail.com
Yo‘ldashev Javohir Rasulbek o‘g‘li
Yuldashevjavohir2405@gmail.com
O‘zMU Jizzax filiali, Amaliy matematika fakulteti talabalari
Xoljigitov Dilmurod Xolmurod o‘g‘li
Ilmiy rahbar, O‘zMU Jizzax filiali katta o’qituvchisi

Annotatsiya: Ushbu ishda kompyuter viruslarini aniqlash va ularni turlarga ajratishda matematik klasterlash algoritmlarining roli o‘rganiladi. Ma’lumotlar asosida viruslarning o‘xshash xususiyatlarini aniqlash uchun “K-means”, “DBSCAN” va “Hierarchical clustering” algoritmlari qo‘llaniladi. Tadqiqotning maqsadi — yangi aniqlanmagan viruslarni ularning xatti-harakati yoki kod tuzilishiga qarab avtomatik ravishda guruhlash orqali “tezkor javob berish tizimini” yaratishdir. Natijada, klasterlash usullari yordamida zararli dasturlarni erta aniqlash, antivirus dasturlarining samaradorligini oshirish va kiberxavfsizlikni mustahkamlash imkoniyati yaratiladi.

Kalit so‘zlar: Kompyuter viruslari, klasterlash algoritmlari, K-means, DBSCAN, Hierarchical clustering, mashinaviy o‘rganish, ma’lumotlarni tahlil qilish, virus klassifikatsiyasi, kiberxavfsizlik, zararli dasturlarni aniqlash.

Kirish: Zamonaviy axborot texnologiyalari rivoji bilan bir qatorda, kiberxavfsizlik masalalari ham dolzarb muammolardan biriga aylandi. Har kuni minglab yangi kompyuter viruslari, zararli dasturlar (malware), troyanlar, fishing dasturlari va boshqa turdagi kiber tahdidlar paydo bo‘lmoqda. Bunday sharoitda an’anaviy antivirus tizimlari — signatura yoki qoidalar asosida ishlovchi himoya mexanizmlari — o‘z samaradorligini tobora yo‘qotmoqda. Sababi, yangi virus turlari mavjud viruslardan farqli xususiyatlarga ega bo‘lib, ular o‘zini yashirish, xatti-harakatini o‘zgartirish yoki turli kriptografik himoya qatlamlaridan foydalanish orqali aniqlanishdan qochadi. Shu bois, zamonaviy kiberxavfsizlik sohasida viruslarni aniqlashda matematik modellash va klasterlash algoritmlaridan foydalanish istiqbolli yo‘nalish hisoblanadi.

Matematik klasterlash usullari – bu katta hajmdagi ma’lumotlar to‘plamida o‘xshash elementlarni guruhlash imkonini beruvchi algoritmlar to‘plamidir. Ularning asosiy maqsadi — oldindan belgilangan yorliqlarsiz (unsupervised learning)

ma’lumotlar orasidagi ichki bog‘liqlik va tuzilmani aniqlashdir. Kompyuter viruslarini tahlil qilishda bu yondashuv ayniqsa foydalidir, chunki ko‘p hollarda viruslar oldindan belgilangan toifalarga kiritilmagan bo‘ladi. Klasterlash algoritmlari esa ularni xatti-harakat naqshlari, kod strukturalari yoki resurslardan foydalanish ko‘rsatkichlari asosida o‘xshashlikka qarab guruhlaydi.

K-means algoritmi ma’lumotlarni belgilangan miqdordagi klasterlarga ajratadi va har bir klasterni markaziy nuqta (centroid) orqali ifodalaydi. U samarali va tez ishlovchi algoritm bo‘lib, katta hajmdagi virus ma’lumotlar bazalarida tezkor tahlilni ta’minlaydi.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) esa zichlik asosida klasterlash usuli bo‘lib, murakkab, chiziqli bo‘lmagan tuzilmalarni aniqlashda samaralidir. Bu algoritm ayniqsa, turli xil virus namunalarining “shovqinli” ma’lumotlar orasida aniqlanishini osonlashtiradi.

Hierarchical clustering (ierarxik klasterlash) esa viruslarni daraxtsimon tuzilma asosida tahlil qiladi va ularning o‘zaro qanchalik o‘xshashligini vizual tarzda ko‘rsatish imkonini beradi.

Mazkur yondashuvlarning kombinatsiyasi yangi noma’lum viruslarni aniqlashda katta afzalliklarga ega. Misol uchun, antivirus ishlab chiquvchi kompaniyalar real vaqt rejimida virus namunalarini tahlil qilib, ularning xatti-harakatlarini kuzatadi. So‘ngra ushbu ma’lumotlar klasterlash algoritmlariga kiritiladi. Agar yangi dastur ma’lum bir klasterdagi viruslarning harakatiga o‘xshash bo‘lsa, tizim uni potensial zararli sifatida belgilaydi. Natijada, tizim yangi virusga tezkor javob berishi, u tarqalmasdan oldin uni bloklashi mumkin.

Klasterlash algoritmlarining amaliy qo‘llanishi keng: ular nafaqat antivirus tizimlarida, balki tarmoq trafigini tahlil qilish, soxta xabarlarni (spam) aniqlash, phishing saytlarni guruhlash, va IoT qurilmalardagi zararli faoliyatni monitoring qilish sohalarida ham ishlatilmoqda. Ayniqsa, mashinaviy o‘rganish (machine learning) va sun’iy intellekt texnologiyalari rivoji bilan, klasterlash usullari kiberxavfsizlik tizimlarining asosiy tayanchi sifatida qaralmoqda.

Viruslarni klasterlash orqali aniqlashning muhim afzalliklaridan biri — yangi turlarni signaturasiz aniqlash imkoniyatidir. Bu, ya’ni, virus haqida hech qanday oldindan ma’lumot bo‘lmasa ham, tizim ularning o‘xshash xatti-harakatini asos qilib, ehtimoliy xavf sifatida aniqlashi mumkin. Natijada, kiberxavfsizlik tizimi ancha “aqlli” bo‘ladi, ya’ni o‘rganish, moslashish va yangi tahdidlarga tezda javob qaytarish qobiliyatiga ega bo‘ladi.

Tahlillar shuni ko‘rsatadiki, K-means algoritmi katta ma’lumotlar bazasida samarali ishlasa, DBSCAN esa murakkab, chiziqli bo‘lmagan ma’lumot turlarida yaxshiroq natija beradi. Hierarchical clustering esa viruslar orasidagi nasabiy o‘xshashlikni ko‘rsatishda eng qulay vositadir. Shu sababli zamonaviy kiberxavfsizlik

tadqiqotlarida ushbu algoritmlarni birgalikda qo‘llash orqali gibrid klasterlash tizimlari ishlab chiqilmoqda.

Tadqiqotlar natijasida aniqlanishicha, klasterlashga asoslangan klassifikatsiya usullari an’anaviy antivirus signatura bazasiga tayanadigan tizimlardan 25–40% gacha samaraliroq bo‘lishi mumkin. Bu usul yordamida yangi noma’lum viruslarni erta bosqichda aniqlash, ularning tarqalishini to‘xtatish va tizimlarni avtomatik ravishda himoya qilish imkoniyati yuzaga keladi.

Shuningdek, matematik klasterlash algoritmlarini yanada mukammallashtirish uchun vaznli atribut tanlash, masofa o‘lchovlarini moslashtirish va avtomatik klaster sonini aniqlash kabi yo‘nalishlarda ilmiy ishlar olib borilmoqda. Bu esa, kompyuter viruslarini klassifikatsiya qilish jarayonini yanada aniqlashtiradi va tahlil tezligini oshiradi.

Xulosa

Kompyuter viruslarini klassifikatsiya qilishda matematik klasterlash algoritmlarining qo‘llanilishi zamonaviy kiberxavfsizlik tizimlarida muhim o‘rin tutadi. Ushbu yondashuvlar yordamida viruslar ularning xatti-harakati, kod tuzilishi va o‘xshashlik darajasi asosida guruhlanadi, bu esa yangi va noma’lum tahdidlarni erta bosqichda aniqlash imkonini beradi. K-means, DBSCAN va Hierarchical clustering algoritmlari bu jarayonda eng samarali usullar hisoblanadi — ular viruslarni ma’lumotlar orasidagi yashirin bog‘liqliklarni tahlil qilib, avtomatik tarzda klasterlarga ajratadi.

Natijada, klasterlashga asoslangan tahlil an’anaviy signatura asosidagi antivirus tizimlariga qaraganda tezkor, moslashuvchan va intellektual himoya mexanizmini yaratish imkonini beradi. Bu usul yordamida antivirus tizimlari inson aralashuvisiz ham yangi tahdidlarni aniqlab, ularga javob qaytara oladi. Shunday qilib, matematik klasterlash algoritmlari kelajakda aqlli antivirus tizimlari va avtomatlashtirilgan xavfsizlik tahlilchilarini yaratishda ilmiy va amaliy jihatdan muhim yo‘nalish sifatida qaraladi.

Foydalangan adabiyotlar ro‘yhati

1. *“Online Clustering of Known and Emerging Malware Families”* — maqolada k-means, DBSCAN va onlayn klasterlash algoritmlarining kombinatsiyasi, oqim ma’lumotlarda yuqori tasiqlik (purity) natijasi haqida ma’lumotlar keltirilgan.
2. https://arxiv.org/pdf/2405.03298
3. *“Classification and Online Clustering of Zero-Day Malware”* — yangi virus namunalari uchun real-vaqt klasterlash va klassifikatsiya modelini taklif qiladi.
4. https://link.springer.com/article/10.1007/s11416-024-00513-5?utm_source
5. *“Malware Classification based on Call Graph Clustering”* — viruslarni chaqiruv graflari orqali klasterlash usuli, graflar orasidagi similarlik va DBSCAN/k-medoids qo‘llanilishi.
6. https://arxiv.org/abs/1008.4365?utm_source

7. *"Clustering Analysis for Malware Behavior Detection using Registry Data"* — Windows reyestri ma'lumotlaridan foydalanib K-means yordamida zararli va oddiy jarayonlarni guruhlash bo'yicha tadqiqot.
8. https://thesai.org/Downloads/Volume10No12/Paper_13-Clustering_Analysis_for_Malware_Behavior_Detection.pdf?utm_source
9. *"Automatic Malware Categorization Based on K-Means Clustering Technique"* — Android muhitida malware'larni K-means yordamida avtomatik kategoriyalash usuli.
10. https://www.researchgate.net/publication/355217172_Automatic_Malware_Categorization_Based_on_K-Means_Clustering_Technique?utm_source
11. *"Partitional Clustering of Malware Using K-Means"* — malware xatti-harakat ma'lumotlarini bo'lish (partition) orqali K-means klasterlash modeli.
12. https://www.researchgate.net/publication/263129107_Partitional_Clustering_of_Malware_Using_K-Means
13. *"Clustering for malware classification"* — klasterlash usullarining malware klassifikatsiyasiga tatbiqi umumiy sharhi.
14. https://www.researchgate.net/publication/292185741_Clustering_for_malware_classification
15. *"Static Malware Family Clustering via Structural and Functional"* — static tahlil atributlari bo'yicha struktura va funksiya ma'lumotlari asosida klasterlash.
16. https://scholar.smu.edu/cgi/viewcontent.cgi?article=1248&context=datasciencereview&utm_source
17. *"Machine Learning Aided Static Malware Analysis: A Survey and Tutorial"* — statik tahlil va mashinaviy o'rganish usullari, shu jumladan klasterlash usullari bo'yicha ko'p manbali sharh.
18. https://arxiv.org/abs/1808.01201?utm_source
19. *"A Survey on Malware Detection with Graph Representation Learning"* — malware'ni graflar ko'rinishida ifodalash va graf asosli usullar (shu jumladan klasterlash / GNN) bo'yicha tendensiyalar va misollar.
20. https://arxiv.org/abs/2303.16004?utm_source