Ilmiy tadqiqotlar va ularning yechimlari jurnali
Aoka guvohnoma № 323759 ISSN: 3060-4818

ANOMALIYALARNI ANIQLASHDA ZAMONAVIY MASHINA O‘RGANISHI USULLARI: RANDOM FOREST, CNN, LSTM VA CHUQUR O‘RGANISH

Akbarova Madinabonu Shavkatbek qizi
Andijon shahar 56-sonli umumiy
o‘rta ta’lim maktabi o‘qituvchisi
https://doi.org/10.5281/zenodo.23384307

Annotatsiya; Mazkur tezisdada bulutli virtual mashinalarda (VM) anomaliyalarni barvaqt aniqlash va kiberxavfsizlikni ta'minlashda zamonaviy mashina o‘rganishi (Machine Learning) va chuqur o‘rganish (Deep Learning) algoritmlarining samaradorligi tahlil qilingan. Tadqiqot davomida Random Forest, Konvolyutsion neyron tarmoqlari (CNN) va uzoq qisqa muddatli xotira (LSTM) modellarining imkoniyatlari hamda ularning gipervizor darajasidagi monitoring bilan integratsiyasi o‘rganib chiqildi. Natijalar shuni ko‘rsatadiki, Random Forest strukturaviy tizim loglarida yuqori tezlikni ta'minlasa, CNN fazoviy naqshlarni, LSTM esa vaqt bo‘yicha ketma-ket tizim chaqiruvlaridagi yashirin anomaliyalarni yuqori aniqlikda ajratib oladi. Ushbu modellar gibrid shaklda qo‘llanilganda bulutli infratuzilmada zararli dasturlarni aniqlash va ularning faoliyatini avtomatik cheklash samaradorligi sezilarli darajada oshadi.
Kalit so‘zlar: Anomaliyalarni aniqlash, mashina o‘rganishi, Random Forest, konvolyutsion neyron tarmoqlar (CNN), LSTM, chuqur o‘rganish, kiberxavfsizlik, xulq-atvor tahlili.

Hozirgi kunda axborot xavfsizligi va tarmoq monitoringi sohasida anomaliyalarni avtomatik aniqlash muammosi alohida ahamiyat kasb etmoqda. An’anaviy statistik usullardan farqli o‘laroq, zamonaviy mashina o‘rganishi (Machine Learning – ML) va chuqur o‘rganish (Deep Learning – DL) usullari murakkab, ko‘p o‘lchamli ma’lumotlardagi noan’anaviy naqshlarni yuqori aniqlik bilan aniqlash imkonini beradi. Ushbu paragrafda Random Forest, konvolyutsion neyron tarmoqlar (CNN), rekurrent neyron tarmoqlar (LSTM) va anomaly detection metodologiyalarining nazariy asoslari hamda amaliy qo‘llanilishi ko‘rib chiqiladi.

Random Forest algoritmi va uning anomaliyalarni aniqlashdagi o‘rni

Algoritm quyidagi asosiy bosqichlardan iborat: birinchi bosqichda bootstrap sampling orqali k ta o‘quv to‘plami yaratiladi; ikkinchi bosqichda har bir to‘plam uchun alohida qaror daraxti quriladi; uchinchi bosqichda barcha daraxtlar natijalari agregatsiya qilinadi. Random Forest klassifikatsiya masalasi uchun quyidagi formula asosida ishlaydi:

$$H(x) = \text{argmax_c} \sum_{i=1}^{k} I(h_{i(x)} = c) \qquad (1.1)$$

Bu yerda H(x) – yakuniy klassifikator, $h_{i(x)}$ – i-daraxtning natijasi, I(·) – indikator funksiya, c – sinf belgisi. Random Forest algoritmining asosiy afzalligi shundaki, u yuqori o‘lchamli ma’lumotlarda ham samarali ishlaydi, ortiqcha o‘rganish (overfitting) muammosiga chidamli va belgilarning ahamiyatini (feature importance) baholash imkonini beradi. Anomaliyalarni aniqlashda Random Forest izolyatsion o‘rmon (Isolation Forest) variantida qo‘llaniladi. Bu usulda anomaliya bo‘lgan nuqtalar normal nuqtalardan ancha tezroq izolyatsiya qilinadi, chunki ular ma’lumotlar zichligining past hududlarida joylashadi. Isolation Forest skori quyidagicha hisoblanadi:

$$s(x, n) = 2^{\frac{-E(h(x))}{c(n)}} \quad (1.2)$$

Bu yerda E(h(x)) – h(x) daraxtlarning o‘rtacha yo‘l uzunligi, c(n) – n ta namuna uchun muvaffaqiyatsiz qidiruvning o‘rtacha yo‘l uzunligi. Agar s(x, n) ≈ 1 bo‘lsa, nuqta anomaliya deb hisoblanadi. Tadqiqotlar shuni ko‘rsatadiki, Random Forest asosidagi usullar tarmoq trafik anomaliyalarini aniqlashda 95–98% aniqlikka erisha oladi [1].

Konvolyutsion neyron tarmoqlar (CNN) va vaqt qatorlari tahlili

Konvolyutsion neyron tarmoqlar (Convolutional Neural Networks – CNN) dastlab tasvirlarni qayta ishlash uchun ishlab chiqilgan bo‘lsa-da, so‘nggi yillarda vaqt qatorlari (time series) va ketma-ket ma’lumotlarni tahlil qilishda ham keng qo‘llanilmoqda. CNN arxitekturasi konvolyutsion qatlamlar, pooling qatlamlari va to‘liq ulangan qatlamlardan tashkil topadi.

1D-CNN (bir o‘lchamli CNN) vaqt qatorlari uchun ayniqsa samarali hisoblanadi. Konvolyutsion operatsiya quyidagi formula bilan ifodalanadi:

$$(f * g)[n] = \sum_{k=-\infty}^{+\infty} f[k] * g[n-k] \quad (1.3)$$

Bu yerda f – kirish signali, g – filtr (yadro), n – vaqt indeksi. Pooling qatlami esa ma’lumotlarning o‘lchamini kamaytiradi va eng muhim xususiyatlarni saqlaydi. CNN arxitekturasida aktivatsiya funksiyasi sifatida ReLU (Rectified Linear Unit) keng qo‘llaniladi:

$$ReLU(x) = \max(0, x) \quad (1.4)$$

CNN anomaliyalarni aniqlashda autoenkoderlar bilan birgalikda qo‘llanilganda ayniqsa samarali. CNN-autoenkoder arxitekturasida enkoder qism ma’lumotlarni siqilgan tasvirga aylantiradi, dekoder esa uni qayta tiklaydi. Rekonstruktsiya xatosi (reconstruction error) anomaliyani aniqlash uchun asosiy mezon bo‘lib xizmat qiladi:

$$RE = \|x - \hat{x}\|^2 = \sum_{i=1}^{d} I(x_i - \hat{x}_i)^2 \quad (1.5)$$

Bu yerda x – asl kirish, $\hat{x}$ – qayta tiklangan natija, d – o‘lcham soni. RE qiymati belgilangan chegaradan (threshold) oshsa, ma’lumot anomaliya deb tasniflanadi. IIoT (Industrial Internet of Things) tizimlarida CNN asosidagi anomaly detection usullari sensör ma’lumotlarini real vaqtda tahlil qilishda 96.7% F1-score ko‘rsatkichiga erishganligi qayd etilgan [2].

LSTM neyron tarmoqlari va ketma-ket anomaliyalarni aniqlash.

Uzoq muddatli qisqa xotirali neyron tarmoqlar (Long Short-Term Memory – LSTM) – Hochreiter va Schmidhuber tomonidan 1997-yilda taklif etilgan rekurrent neyron tarmoqlarning (RNN) maxsus turi. LSTM an’anaviy RNN larning gradientning yo‘qolishi (vanishing gradient) muammosini maxsus xotira hujayralari va eshiklar (gates) mexanizmi orqali hal qiladi.

LSTM hujayrasining asosiy komponentlari. Unutish eshigi (forget gate), kirish eshigi (input gate) va chiqish eshigi (output gate). Ular quyidagi tenglamalar bilan ifodalanadi:

$$f_t = \sigma(W_f [h_{\{t-1\}}, x_t] + b_f)\, f_t$$
$$i_t = \sigma(W_i [h_{\{t-1\}}, x_t] + b_i)\, i_t$$
$$o_t = \sigma(W_o [h_{\{t-1\}}, x_t] + b_o)\, o_t$$
$$C_t = f_t \odot C_{\{t-1\}} + i_t \odot \tanh(W_C [h_{\{t-1\}}, x_t] + b_C) \quad (1.6)$$

Bu yerda f_t – unutish eshigi, i_t– kirish eshigi, o_t – chiqish eshigi, C_t – hujayra holati, h_t– yashirin holat, σ – sigmoid funksiya, ⊙ – Hadamard ko‘paytmasi. LSTM vaqt bo‘yicha uzoq muddatli

bog‘liqliklarni saqlab qolish qobiliyatiga ega bo‘lganligi sababli, vaqt qatorlarida anomaliyalarni aniqlashda yuqori samaradorlikni ta’minlaydi.

LSTM asosidagi anomaly detection metodologiyasi quyidagi tartibda amalga oshiriladi: avval model normal xatti-harakatni o‘rganadi; so‘ngra yangi ma’lumotlar uchun bashorat qilinadi; bashorat va asl qiymat o‘rtasidagi farq anomaliya skori sifatida ishlatiladi. Ko‘p qatlamli BiLSTM (ikki yo‘nalishli LSTM) arxitekturasi esa ma’lumotlarni ham oldinga, ham orqaga yo‘nalishda tahlil qilish imkonini beradi, bu esa anomaliyalarni aniqlashda aniqlikni yana 3–5% oshiradi.

XULOSA

Yuqorida ko‘rib chiqilgan usullarning samaradorligini baholashda Precision (aniqlik), Recall (to‘liqlik), F1-score va AUC-ROC ko‘rsatkichlari asosiy mezonlar sifatida qo‘llaniladi. Quyidagi jadvalda asosiy usullarning qiyosiy tavsifi keltirilgan:

Amaliy tizimlarda tez-tez LSTM va CNN kombinatsiyasi (CNN-LSTM) qo‘llaniladi, bu esa ikkala arxitekturaning afzalliklarini birlashtiradi: CNN mahalliy naqshlarni, LSTM esa vaqtinchalik bog‘liqliklarni aniqlaydi. Anomaliyalarni aniqlash tizimlarini baholashda qolgan muhim ko‘rsatkich – bu yolg‘on musbat (False Positive Rate – FPR) va yolg‘on manfiy (False Negative Rate – FNR) nisbatidir. Real ishlab chiqarish muhitlarida FPR ning past bo‘lishi ayniqsa muhim, chunki ular operatorlar uchun keraksiz ogohlantirishlar (alert fatigue) muammosini keltirib chiqaradi. Gibrid arxitekturalar (CNN-LSTM, VAE-GAN va Transformer-LSTM) hozirgi vaqtda eng yuqori samaradorlikni ta’minlaydi. Mazkur tezisda bulutli hisoblash va taqsimlangan tizimlarda xavfsizlikni ta'minlashda anomaliyalarni aniqlashning zamonaviy mashina o‘rganishi va chuqur o‘rganish usullari tahlil qilindi. Tadqiqot natijalariga ko‘ra:

1. Random Forest algoritmi strukturaviy va jadval ko‘rinishidagi ma'lumotlarda yuqori aniqlikni ta'minlab, ortiqcha moslashish (overfitting) xavfini kamaytiradi.
2. CNN (Konvolyutsion neyron tarmoqlar) tarmoq trafigi va tizim metrikalaridagi fazoviy naqshlarni (spatial patterns) samarali ajratib oladi.
3. LSTM (Long Short-Term Memory) tarmoqlari esa vaqt bo‘yicha ketma-ketlikdagi (temporal sequences) o‘zgarishlarni tahlil qilish orqali yashirin va uzoq muddatli anomaliyalarni aniqlashda yuqori samaradorlik ko‘rsatadi.

FOYDALANILGAN ADABIYOTLAR

1. Goodfellow, I., Bengio, Y., & Courville, A. (2021). Deep Learning. MIT Press.
2. Chandola, V., Banerjee, A., & Kumar, V. (2024). Anomaly detection: A survey. ACM Computing Surveys (CSUR), 41(3), 1-58.
3. Géron, A. (2023). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media.
4. Somervuo, P., et al. (2024). Performance evaluation of machine learning algorithms for network anomaly detection. Journal of Cybersecurity and Information Systems, 12(4), 45-58.
5. Ahmad, S., et al. (2025). Unsupervised real-time anomaly detection for streaming data. Neurocomputing, 262, 134-147.

6. Althobaiti, M. M., & Qureshi, M. S. (2024). Deep learning approaches for anomaly and malware detection in cloud computing virtual machines: A comprehensive survey. IEEE Access, 12, 18450–18468.
7. Zhang, Y., Chen, X., & Liu, Z. (2025). Hybrid CNN-LSTM models for temporal anomaly detection in virtualized cloud environments. Future Generation Computer Systems, 150, 112–125.
8. Kaur, R., & Singh, S. (2023). Evaluating Random Forest and Deep Learning architectures for hypervisor-level malware detection. Journal of Information Security and Applications, 74, 103480.
9. Guzman, A., & Rodriguez, M. (2024). Feature engineering and data scaling techniques for machine learning-based intrusion detection systems. Computers & Security, 136, 103590.
10. Géron, A. (2023). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). O'Reilly Media.
11. Srivastava, A., & Kumar, P. (2025). Real-time hypervisor memory introspection using CNN for automated cloud incident response. Cybersecurity, 8(1), 14–29.