ЛЕКЦИЯ 14
Кибершабуылдарды интеллектуалды анықтау жүйелерінің тиімділігін бағалау

Лекция мақсаты

Интеллектуалды кибершабуыл детекторларын дұрыс бағалау қағидаларын түсіндіру; confusion matrix, accuracy, precision, recall, F1, ROC-AUC, жалған іске қосылу, деректерді бөлу, қайталанғыштық және операциялық көрсеткіштерді талдау.

Тиімділікті бағалаудың мақсаты

Интеллектуалды анықтау жүйесін бағалау — модельдің тест дерегінде жоғары сан көрсетуін ғана емес, нақты қауіпсіздік міндетін қаншалықты дұрыс орындайтынын анықтау. Бағалау деректер сапасын, класстар теңгерімін, қате оң және қате теріс шешімдерді, анықтау уақытын және операциялық әсерді қамтиды.

Мысал. Шабуылдар сирек кездесетін деректерде барлық жазбаны «қалыпты» деп белгілеген модель жалпы accuracy бойынша жоғары көрінуі мүмкін, бірақ қауіпсіздік тұрғысынан пайдасыз.

Confusion matrix

Confusion matrix модель шешімін төрт жағдайға бөледі: True Positive — шабуылды дұрыс анықтау; True Negative — қалыпты әрекетті дұрыс тану; False Positive — қалыпты әрекетті шабуыл деп белгілеу; False Negative — шабуылды өткізіп алу.

FP аналитик жүктемесін арттырады және қажетсіз блоктауға әкелуі мүмкін. FN нақты шабуылдың байқалмай қалуына әкеледі. Қай қате маңыздырақ екені қорғалатын актив пен қолдану сценарийіне тәуелді.

Accuracy, Precision, Recall және F1

Accuracy — барлық шешім ішіндегі дұрыс шешім үлесі. Precision — модель шабуыл деп белгілеген жағдайлардың қаншасы шынында шабуыл болғанын көрсетеді. Recall — нақты шабуылдардың қаншасы анықталғанын көрсетеді. F1 precision мен recall арасындағы теңгерімді сипаттайды.

Мысал. SOC-та мыңдаған жалған ескерту аналитикті шаршатса, precision маңызды болады. Ал аса маңызды жүйеде шабуылды өткізіп алу өте қауіпті болса, recall-ға ерекше назар аударылады. Бір ғана метрика барлық жағдайға жеткіліксіз.

ROC-AUC және шекті мән

Көп модель ықтималдық немесе score береді, ал соңғы шешім threshold арқылы жасалады. Шекті мән төмендесе көбірек оқиға шабуыл деп белгіленеді: recall өсуі мүмкін, бірақ false positive те көбейеді. Шекті мән жоғарыласа керісінше әсер болуы ықтимал.

ROC қисығы әртүрлі threshold жағдайында true positive rate пен false positive rate арақатынасын көрсетеді, ROC-AUC модельдің класстарды ажырату қабілетін жинақтап сипаттайды. Бірақ сирек шабуылдар бар ортада precision-recall көрсеткіштерін де қарастыру маңызды.

Деректерді бөлу және leakage

Train, validation және test жиындары тәуелсіз бағалауға мүмкіндік беруі тиіс. Data leakage — тест туралы ақпарат оқытуға тікелей немесе жанама өтіп кеткен жағдай. Киберқауіпсіздік деректерінде бір сессияның немесе бір шабуыл кампаниясының өте ұқсас жазбаларын train мен test-ке кездейсоқ бөлу жасанды жоғары нәтиже беруі мүмкін.

Мысал. Бір желілік сессиядан алынған пакеттердің бір бөлігі оқытуға, қалғаны тестке түссе, модель шабуылдың жалпы заңдылығын емес, сол сессияның ерекшелігін жаттап алуы мүмкін. Сондықтан бөлу сессия, уақыт немесе субъект деңгейінде ұйымдастырылуы мүмкін.

Тұрақтылық және қайталанғыштық

Бір рет алынған нәтиже модель сапасын толық сипаттамайды. Әртүрлі random seed, бірнеше қайталау, орташа мән және стандарттық ауытқу нәтижелердің тұрақтылығын көрсетеді. Гиперпараметрлер, деректерді алдын ала өңдеу және бөлу процедурасы құжатталуы тиіс.

Мысал. Бір іске қосуда F1=0.96, ал басқа іске қосуларда 0.82–0.90 болса, тек ең жоғары санды көрсету модельдің нақты тұрақтылығын жасырған болар еді.

Операциялық метрикалар

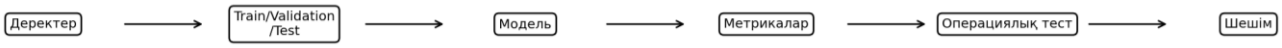
Киберқорғаныс жүйесі үшін модельдік метрикамен қатар анықтау кідірісі, бір секундта өңделетін оқиғалар саны, ресурстық шығын, alert көлемі және әрекет ету уақыты бағаланады. Өте дәл, бірақ оқиғаны бірнеше сағат кеш анықтайтын модель жедел қорғаныс міндетіне сәйкес келмеуі мүмкін.

Мысал. Екі модельдің F1 көрсеткіші ұқсас болса, бірақ біреуі нақты уақыт ағынын өңдей алса, ал екіншісі тек офлайн режимде жұмыс істесе, олардың практикалық құндылығы әртүрлі.

Негізгі жіктеу және салыстыру

Көрсеткіш	**Мағынасы**	**Қауіпсіздік тұрғысынан түсіндіру**
Accuracy	Жалпы дұрыс шешім үлесі	Теңгерімсіз класта жаңылыстыруы мүмкін
Precision	TP/(TP+FP)	Alert-тердің қаншасы негізді
Recall	TP/(TP+FN)	Нақты шабуылдардың қаншасы табылды
F1	Precision мен Recall үйлесімі	Екі көрсеткіштің теңгерімі
ROC-AUC	Класстарды ажырату қабілеті	Threshold-тан тәуелсіз салыстыру
Detection latency	Анықтауға кеткен уақыт	Жедел әрекет ету мүмкіндігі

Өзара байланыстар схемасы



Сурет 1. Интеллектуалды детекторды бағалау кезеңдері

Кешенді практикалық мысал

Желі аномалияларын анықтайтын модель 99% accuracy көрсетті делік. Деректердің 99%-ы қалыпты, 1%-ы шабуыл болса, бұл сан өздігінен модельдің пайдалы екенін дәлелдемейді. Confusion matrix арқылы қанша шабуыл өткізілгені және қанша қалыпты оқиға қате белгіленгені тексеріледі.

Кейін train/test бөлінуі зерттеледі. Бір сессияға тиесілі өте ұқсас жазбалар екі жиынға қатар түспеуі керек. Бірнеше random seed бойынша нәтиже қайталанып, орташа көрсеткіш пен өзгергіштік есептеледі. Соңында модель нақты оқиға ағынында тексеріліп, секундына өңдеу қабілеті, alert саны және анықтау кідірісі бағаланады.

Бағалау деректерін қалыптастыру және класстар теңгерімі

Интеллектуалды анықтау жүйесінің нәтижесі бастапқы деректердің сапасына тікелей тәуелді. Бағалау жиынтығында қалыпты әрекеттер мен шабуыл үлгілері нақты жұмыс ортасына жақын арақатынаста болуы керек. Егер бір класс жасанды түрде шамадан тыс көбейтілсе немесе өте аз ұсынылса, алынған метрикаларды нақты эксплуатациялық ортаға тікелей көшіруге болмайды. Сондықтан деректердің қайдан алынғаны, қандай уақыт аралығын қамтитыны, қай шабуыл түрлері бар екені және қай объектілерден жиналғаны міндетті түрде сипатталады.

Мысал. Желілік трафик деректерінде 100 000 қалыпты жазба және 1 000 шабуыл жазбасы бар делік. Мұндай жағдайда класстар теңгерімсіз. Модель 99 000 қалыпты жазбаны дұрыс танып, бірақ шабуылдардың едәуір бөлігін өткізіп алса, жалпы accuracy әлі де жоғары көрінуі мүмкін. Сондықтан confusion matrix, precision және recall бірге қарастырылады.

Precision мен Recall арасындағы таңдау

Precision және Recall бір-бірімен байланысты, бірақ әртүрлі операциялық сұраққа жауап береді. Precision жоғары болса, модель берген ескертулердің көп бөлігі негізді болады. Recall жоғары болса, нақты шабуылдардың көп бөлігі анықталады. Threshold өзгерген кезде бұл көрсеткіштердің арақатынасы да өзгереді. Сондықтан шекті мәнді тек математикалық максимум бойынша емес, SOC мүмкіндігі мен қорғалатын активтің маңыздылығы бойынша таңдау қажет.

Мысал. Егер жүйе тәулігіне миллиондаған оқиға өңдесе, өте төмен threshold мыңдаған қосымша alert туғызуы мүмкін. Аналитиктер оларды өңдей алмаса, жоғары recall-дың практикалық пайдасы төмендейді. Ал аса маңызды инфрақұрылымда сирек, бірақ қауіпті шабуылды өткізіп алу қолайсыз болуы мүмкін; бұл жағдайда recall басымдық алады.

PR-қисық және теңгерімсіз деректер

Шабуыл оқиғалары қалыпты оқиғаларға қарағанда сирек болған кезде Precision–Recall қисығы модель сапасын түсіндіруге пайдалы. Ол threshold өзгерген сайын precision мен recall қалай өзгеретінін көрсетеді. Бұл әсіресе «оң класс» ретінде шабуыл сирек кездесетін жағдайда маңызды, себебі false positive саны аз пайыз болғанның өзінде абсолюттік мәнде өте үлкен болуы мүмкін.

Модельдерді әділ салыстыру

Екі модельді салыстырғанда олар бірдей train, validation және test бөлінуінде, бірдей алдын ала өңдеу шарттарында және бірдей метрика есептеу процедурасында бағалануы тиіс. Бір модельге тиімді параметрлерді ұзақ баптап, екіншісін әдепкі параметрлермен қалдыру әділ салыстыру болмайды. Гиперпараметрлерді таңдау test жиынтығына қарап жасалмауы керек; test тек соңғы тәуелсіз бағалау үшін сақталады.

Мысал. Бір зерттеуде Random Forest, SVM және нейрондық желі салыстырылсын. Әр модельдің гиперпараметрлері validation жиынтығында таңдалып, соңғы нәтиже бірдей test жиынтығында есептелуі керек. Сонда айырмашылық алгоритмнің нақты қабілетін жақсырақ көрсетеді.

Қате оң және қате теріс нәтижелердің құны

Киберқауіпсіздікте FP және FN тек статистикалық қате емес, нақты операциялық шығынға ие. False Positive аналитик уақытын алады, пайдаланушыны қажетсіз бұғаттауы немесе бизнес-процесті тоқтатуы мүмкін. False Negative шабуылдың байқалмай қалуына, деректердің жоғалуына немесе қарсыластың жүйеде ұзақ уақыт қалуына әкелуі ықтимал. Сондықтан метрикаларды активтің құндылығы және ықтимал салдармен байланыстыру қажет.

Қате түрі	**Не болды?**	**Ықтимал салдар**	**Бағалау кезінде не ескеріледі?**
False Positive	Қалыпты оқиға шабуыл деп белгіленді	Артық alert, қажетсіз блоктау, аналитик жүктемесі	Precision, FPR, alert саны
False Negative	Шабуыл қалыпты деп танылды	Қауіптің байқалмай қалуы	Recall, FNR, актив маңыздылығы
True Positive	Шабуыл дұрыс анықталды	Уақытылы әрекет ету мүмкіндігі	Detection latency, дәлел сапасы
True Negative	Қалыпты әрекет дұрыс танылды	Қажетсіз араласу болмайды	Жүйенің тұрақты жұмысы

Уақыт бойынша валидация және модель дрейфі

Киберорта уақыт өте өзгереді: бағдарламалар жаңартылады, пайдаланушы мінез-құлқы өзгереді, жаңа сервистер қосылады және шабуыл тәсілдері дамиды. Сондықтан кездейсоқ бөлінген

test жиынтығымен қатар уақыт бойынша кейінгі кезеңде жиналған деректерде тексеру пайдалы. Егер модельдің сапасы уақыт өте төмендесе, бұл data drift немесе concept drift белгісі болуы мүмкін.

Мысал. Модель қаңтар–маусым трафигінде оқытылып, шілде–тамыз деректерінде тексеріледі. Егер жаңа VPN сервисі енгізілгеннен кейін false positive күрт өссе, модель жаңа қалыпты мінез-құлықты аномалия деп қабылдап жатқан болуы мүмкін. Мұндай жағдайда модельді қайта оқыту немесе белгілерді қайта қарау қажет.

Түсіндірілетіндік және аналитик сенімі

Интеллектуалды детектордың бағалауына түсіндірілетіндік те кіреді. Аналитикке тек 0,93 тәуекел score-ы емес, қандай белгілер шешімге әсер еткені қажет. Бұл қате оң нәтижені жылдам тануға, инцидентті тексеруге және модельдің күтпеген тәуелділіктерін анықтауға көмектеседі. Түсіндірме модельдің дұрыстығын автоматты түрде дәлелдемейді, бірақ операциялық қолдануды жеңілдетеді.

Бағалаудың толық циклі

Дұрыс бағалау келесі логикамен жүргізіледі: деректердің шығу тегін сипаттау → train/validation/test бөлу → leakage-ті болдырмау → гиперпараметрлерді validation арқылы таңдау → бірнеше seed бойынша қайталау → confusion matrix және негізгі метрикаларды есептеу → threshold таңдау → уақыт пен өнімділік тесті → нақты ортадағы пилоттық тексеру → мониторинг және қайта бағалау. Осы цикл модельдің тек зертханалық емес, эксплуатациялық жарамдылығын бағалауға мүмкіндік береді.

Қорытынды

Интеллектуалды киберқорғаныс жүйесінің сапасын бір ғана accuracy немесе ROC-AUC санымен бағалау жеткіліксіз. Модельдік метрикалар деректерді бөлу дұрыстығымен, қате түрлерінің салдарымен, қайталанғыштықпен және нақты операциялық ортадағы жылдамдық пен жүктememен бірге қарастырылуы тиіс.

Бақылау сұрақтары

1. Confusion matrix қандай төрт жағдайдан тұрады?
2. False Positive қандай салдарға әкеледі?
3. False Negative неге қауіпті?
4. Accuracy қашан жаңылыстыруы мүмкін?
5. Precision мен Recall айырмашылығы қандай?
6. F1 нені көрсетеді?
7. Threshold өзгергенде не өзгереді?
8. Data leakage дегеніміз не?
9. Нәтижені бірнеше seed-пен қайталау не үшін керек?
10. Операциялық метрикаларға не жатады?

Әдебиеттер тізімі:

1. Менщиков А.А., Заколдаев Д.А., Воробьева А.А. Введение в кибербезопасность и киберустойчивость. — СПб.: Университет ИТМО, 2024.

2. Лукьянец С.Е. Гибридные системы кибербезопасности: искусственный интеллект и анализ сетевого трафика. — 2025.