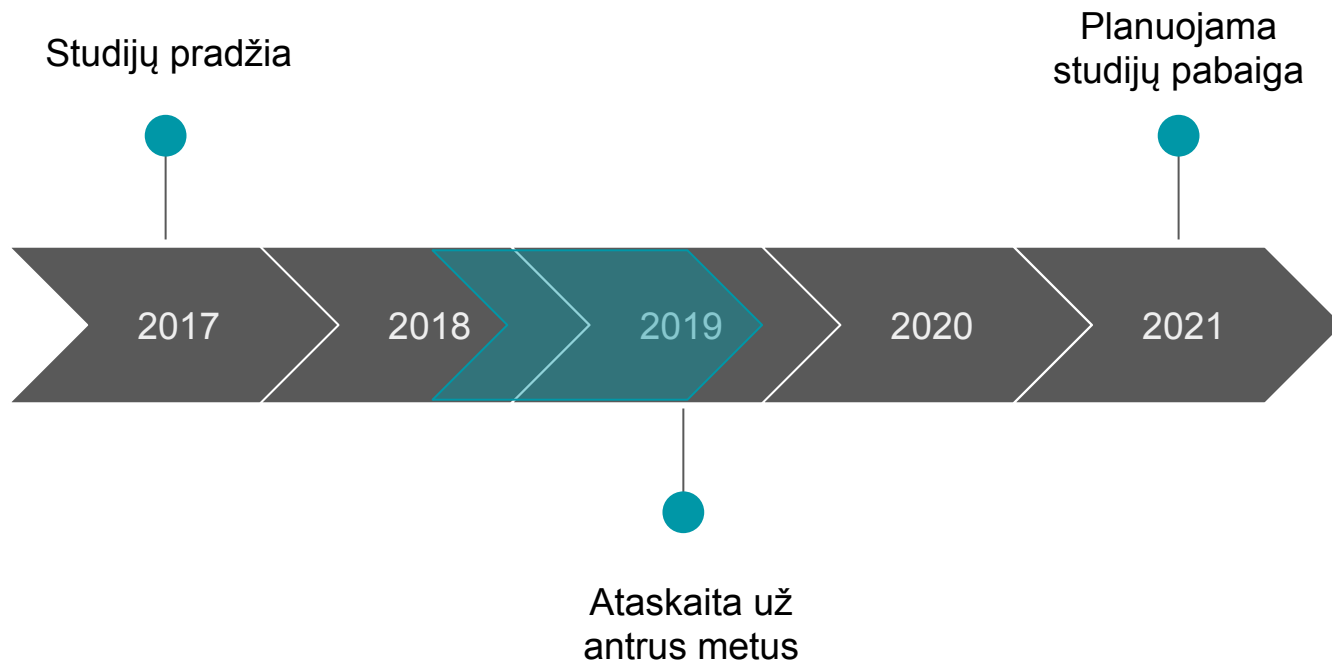


Europos branduolinių mokslinių tyrimų organizacijos (CERN) kompaktiško miuonų solenoido (CMS) eksperimento duomenų sertifikavimas pasitelkiant mašininio mokymo metodus

Doktorantas: Mantas Stankevičius
Vadovas: Dr. Virginijus Marcinkevičius
Konsultantas: Dr. Valdas Rapševičius

Studijos



Tikslas

Sukurti klasifikavimo metodus skirtus CERN CMS duomenų sertifikavimui naudojant mašininį mokymąsi

Objektas

- CERN CMS duomenys
- Klasifikavimo algoritmai
- Anomalijų paieškos algoritmai
- Mašininis mokymasis

Ataskaitinių metų planas

- ✓ Išlaikyti 2 dalykų egzaminus
- ✓ Publikuoti straipsnį
- ✓ Dalyvauti tarptautinėje konferencijoje

Ataskaitinių metų rezultatai

Išlaikyti 2 egzaminai:

- Atpažinimo teorija (10)
- Optimizavimo teorija (7)

Ataskaitinių metų rezultatai

Atlikta neuroninio tinklo hyper parametrų paieška naudojant meta mokymą. Tyrimo rezultatai pateikti straipsnyje:

*Meta-Learning for Artificial Neural Network
Hyper-Parameter Optimization for CERN CMS
Offline Data Certification*

*Straipsnis dar recenzuojamas ir bus publikuotas
Journal Of Physics: Conference Series žurnale*

Ataskaitinių metų rezultatai

Stendinis pranešimas tarptautinėje
konferencijoje *ACAT 2019, Šveicarija*

*Meta-Learning for Artificial Neural Network
Hyper-Parameter Optimization for CERN CMS
Offline Data Certification*

Ataskaitinių metų rezultatai

Dalyvavau vasaros mokykloje liepos
1-5 d. *International Summer School
on Deep Learning 2019, Gdansk,
Lenkija*

Atlikti tyrimai ir pasiekti rezultatai

Data Quality Monitoring

Neapdorotų duomenų ir detektoriaus statuso srautas **nuolat** yra rašomas į duomenų saugyklą

Maža dalis duomenų yra rekonstruojama ir stebima realiu laiku.

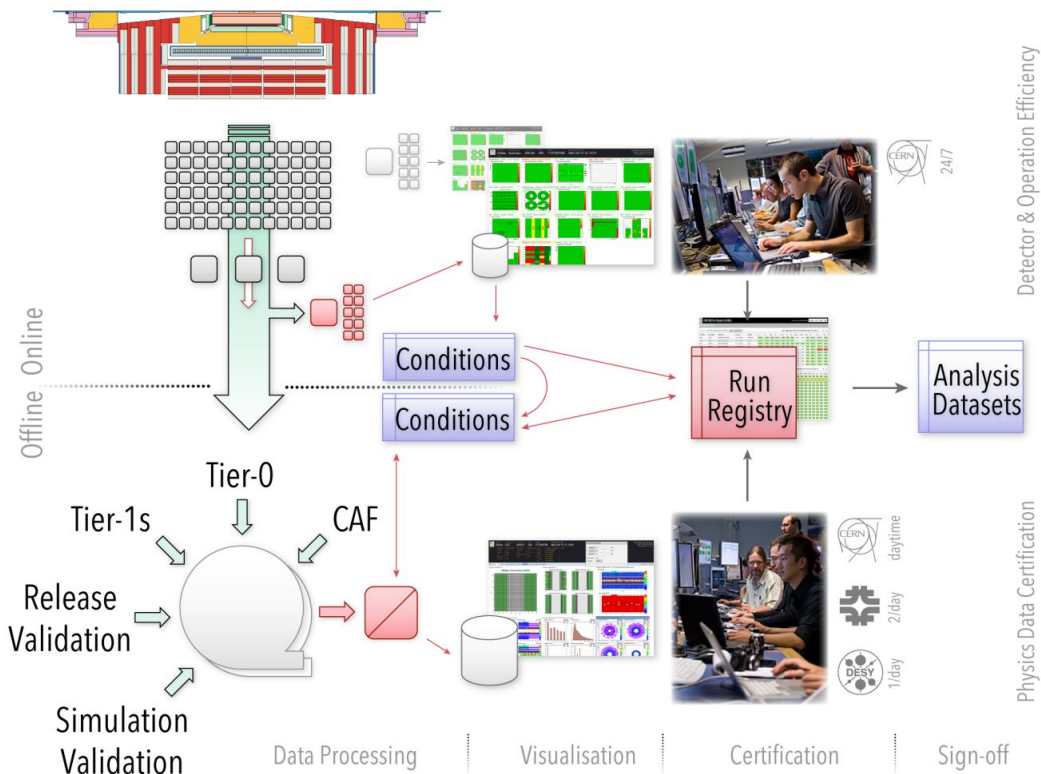
Online shifters pažymi **Run** kaip **GERAS** jeigu yra pakankamai naudingų duomenų ir įranga veikė puikiai

Duomenys yra pilnai rekonstruojami ir kalibruojami **po maždaug 48 valandų**

Offline shifters patikrina svarbiausias histogramas ir nustato duomenų gerumą

Sertifikavimas yra daromas **Run** ir **Lumisekcijų** lygyje

GoldenJSON yra sudaromas tik iš gerų **Run** ir **Lumisekcijų**



Uždavinys

Pirmais metais buvo atliktas klasifikatorių palyginimas. Neuroniniai tinklai nedavė norimo klasifikavimo tikslumo, todėl buvo nuspręsta **tęsti tyrimą** atliekant neuroninio tinklo **hyper parametrų paiešką** taip siekiant padidinti klasifikavimo tikslumą.

	AUC	ACC	F₁
XGB	0.987	0.997	0.998
Random Forest	0.970	0.980	0.990
ANN	0.954	0.961	0.79
Naive Bayes	0.706	0.971	0.985
	Vidurkiai		

Duomenys

Naudojami CERN CMS eksperimento 2016 metų duomenys

- 160.000 lumisekcijų
- 2807 požymiai
- 2 išbalansuotos klasės (santykis 98:2)

Paruošimas:

- Požymių normalizavimas
- Klasių reikšmės imamos iš GoldenJSON

Įranga

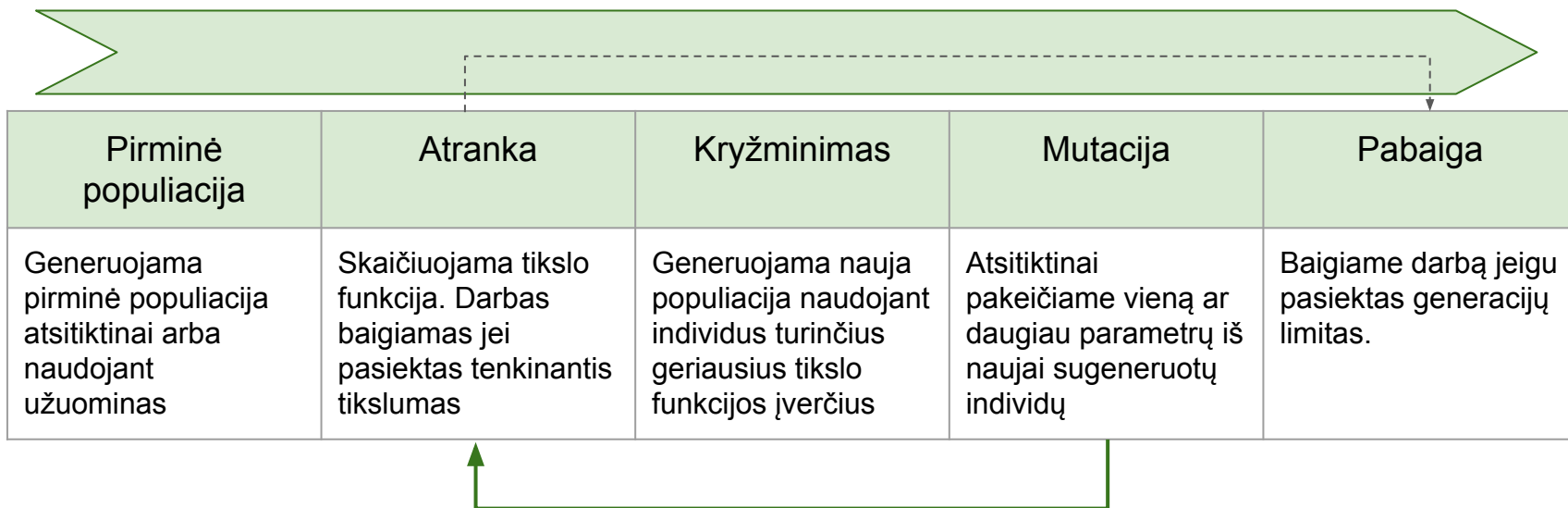
Programinė įranga:

- Python (v3.6)
- Keras (v2.1.5)
- Tensorflow (v1.7)
- scikit-learn (v0.19.1)

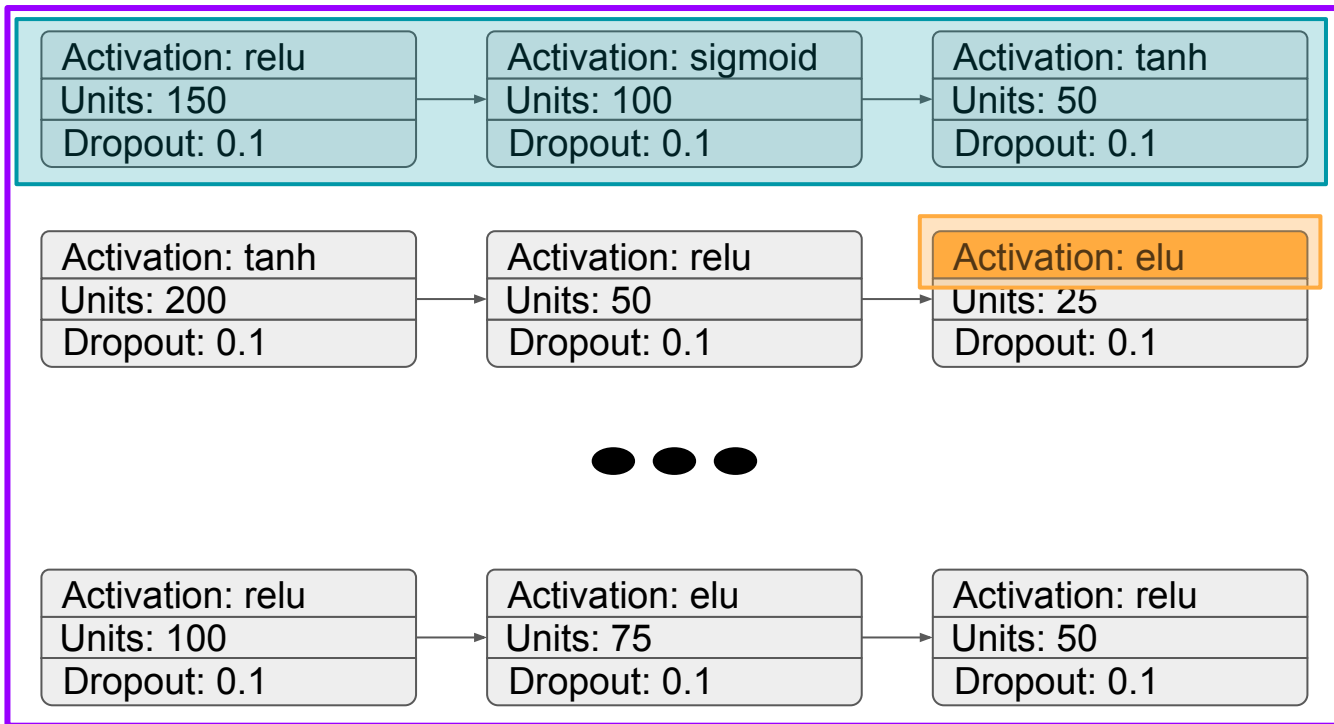
Kompiuterinė įranga:

- PC + NVIDIA GPU (GeForce GTX 1080 Ti)
- VM (8 cores 2.2 GHz, 16GB RAM)

Genetinio algoritmo žingsniai



Populiacija, individas, genai



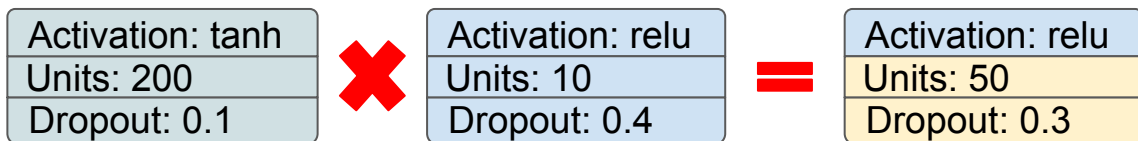
Neuroninis tinklas
(individas)

Sluoksnių parametrai
(genai)

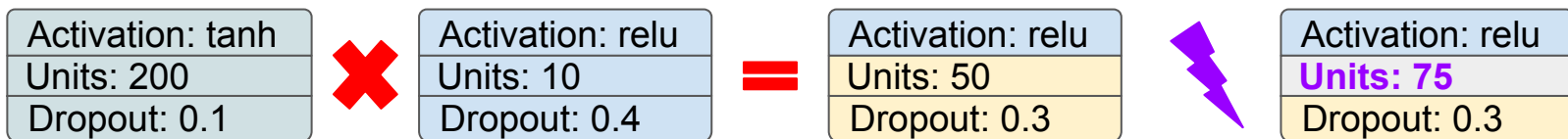
Populiacija

Genetiniai operatoriai

Kryžminimas



Mutacija



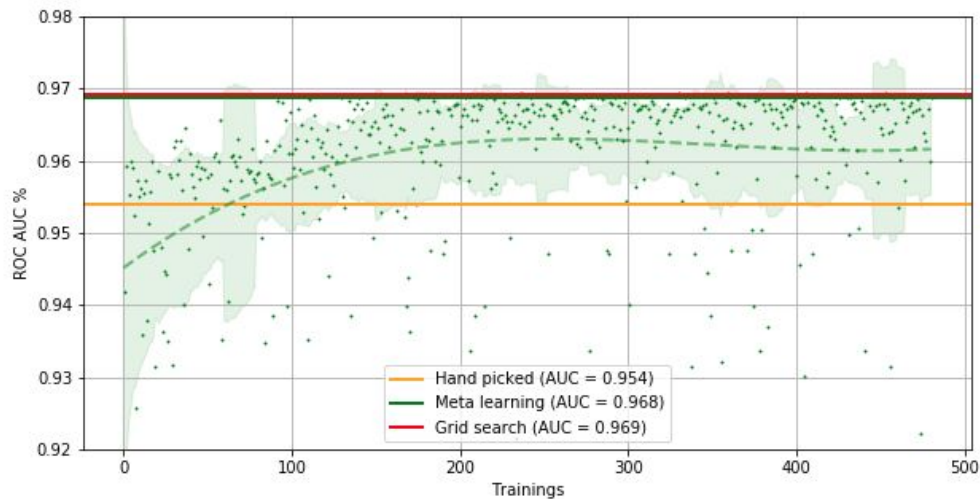
Hyper-Parametrai

Tinklo parametrai	
Neuronų kiekis	[10, 50, 100, 200, 500, 750, 1000, 1500, 2000]
Aktyvacijos funkcija	[relu, elu, sigmoid, tanh]
Dropout	[0, 0.1, 0.2, 0.4, 0.5]
Optimizatorius	[adam, rmsprop, sgd, adagrad, adadelata, adamax]
Genetic algoritmo parametrai	
Populiacijos dydis	50
Generacijų kiekis	∞ (kol gerėja)
Mutacijos dažnis	0.1 (10%)

Rezultatai

Antros evoliucijos metu pasiekiamas tikslumas gautą rankiniu būdu.

Ketvirtos-penktos evoliucijos metu pasiekiamas tikslumas artimas gardelės paieškai.



Metodas	AUC	\pm	Mokymai	Optimizatorius	Tinklas
Rankinis	0.954	0.005	--	Adam	3x [Relu (200) DO (0.5)]
Grid search	0.969	0.002	1080 (4.5d)	Adagrad	Sigmoid (2000) DO (0.2)
Meta learning	0.968	0.002	200 (20h)	Adagrad	Sigmoid (1000) DO (0.2)

Išvados

Naudojant meta mokymą tikėtina rasti optimalų sprendimą greičiau negu perrenkant visus įmanomus variantus.

Šiam uždaviniui pavyko pagreitinti parametrų paiešką **5 kartus**. Hyper parametrų paieška sutrumpėjo nuo **4.5 dienų** iki **20 valandų**.

Meta mokymas gali būti pritaikytas įvairių mašininio mokymosi metodų hyper parametrų paieškai.

Kitų metų planas

- Anomalijų CERN CMS duomenyse analizė
- Mašininio mokymo ir gilaus mokymo metodų naudojamų anomalijų CERN CMS duomenyse aptikimui analizė ir taikymas
- Gautus rezultatus publikuoti recenzuojamame žurnale
- Dalyvauti tarptautinėje konferencijoje

Ačiū

Už dėmesį

