

KUIDAS TEHISARU OSAKESTEFÜÜSIKA MÕÕTMISI TÄPSEMAKS JA EFEKTIIVSEMAKS TEEB

Joosep Pata (Keemilise ja Bioloogilise Füüsika Instituut)

*Teaduspreemia loodusteaduste valdkonnas tööde tsükli
„Masinõppel põhinev suurandmete rekonstrueerimine
osakestefüüsikas“ eest*

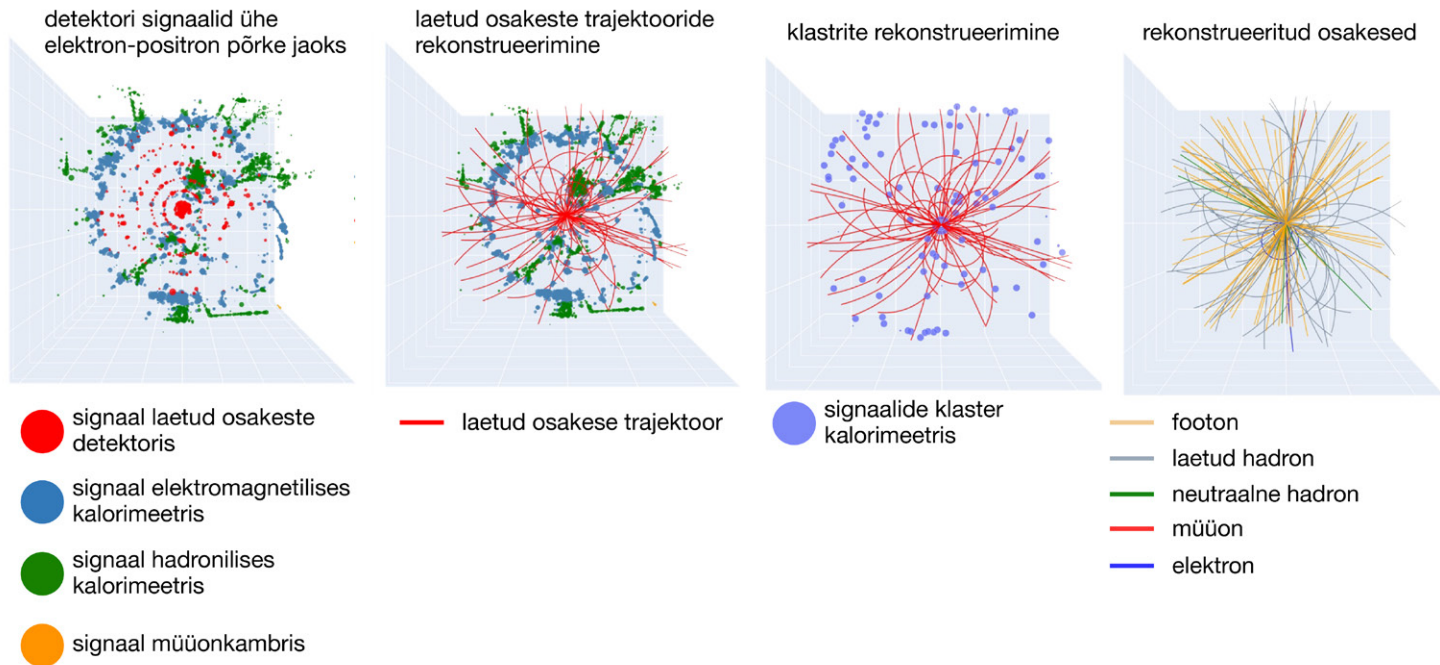
Mis on Higgsi boson ja miks see oluline on?

Kuidas töötab röntgenaparaat või tuumareaktor, millistest tükkidest koosneb Universum ja kuidas need tükid omavahel seostuvad? Kas on veel avastamata loodusjõude ja kuidas need meie elu võivad mõjutada? Nende küsimuste uurimisega tegeleb osakestefüüsika ehk kõrge energia füüsika. Osakestefüüsika on katseline teadus, mis tähendab, et küsimustele tulevad vastused pärismaailmast ehk katsetest. Selline katseaparaat on ehitatud Euroopasse, Šveitsi-Prantsuse piiril asuvasse teaduskeskusse CERN, kus asub maailma suurim ja võimsaim osakeste kiirendi Large Hadron Collider (LHC). Selle aparaadiga avastati aastal 2012 Higgsi boson, täiesti uut tüüpi algosake, mis erineb senimaani avastatud osakestest.

Osakestefüüsikas on praegu üheks peamiseks küsimuseks Higgsi bosoni ja teiste algosakeste vaheliste vastastikmõjude mõõtmine. Higgsi boson on senimaani ainus fundamentaalselt ilma spinnita osake, mis hüpoteesi kohaselt annab läbi Higgsi mehhanismi massi teistele osakestele. Kuivõrd tegu on täiesti uut tüüpi osakesega, on seda mehhanismi tarvis katseliselt täielikult kontrollida, st mõõta Higgsi bosoni vastastikmõju massiivsete osakestega ning iseendaga. Higgsi bosoni vastastikmõju iseendaga ehk eneseinteraktsiooni määramine on CERN-i suure hadronpõrguti LHC üks põhieesmärkidest, kuna see kannab ka informatsiooni Universumi varase arenguetaapi kohta ja võimaldab seeläbi katseliselt uurida tumeainet ja tumeenergiat, kuna nad on jätnud jälje Universumi varases etapis toimunud kiiresse paisumisse ehk inflatsiooni.

Kuidas kasutatakse CERN-i kiirendeid osakeste uurimiseks?

Higgsi bosoni eneseinteraktsiooni mõõtmiseks on kavandatud CERN-i LHC kiirendi kõrghedusfaasi uuendus, kus prootonkiirte pörkes ehk ühes sündmu-



Joonis 1. Osakeste rekonstrueerimise sammud, alustades detektoris registreeritud signaalidest vasakul kuni rekonstrueeritud osakesteni paremal. Andmed on visualiseeritud ühe elektron-positron põrke jaoks tulevikukiirendis CLIC.

ses toimunud interaktsioonide arvu ehk kiire heledust suurendatakse kuni viis korda, et koguda võimalikult palju andmeid LHC eluea jooksul. Heledusfaasi uuendus käivitub plaani järgi aastal 2030, seega käib juba praegu kibekiire töö selle ettevalmistamiseks.

See tähendab, et ühe prootonpõrke käigus võib tekkida kuni sadu tuhandeid osakesi. Nende osakeste omadusi, st trajektoore ja laenguid, mõõdetakse mitmekihiliste detektoritega, mis meenutavad kolmemõõtmelisi kaameraid. Detektorisse tekkinud keerukast kolmemõõtmelisest pildist on vaja korraga rekonstrueerida võimalikult paljude osakeste omadused. Seda probleemi on võimalik käsitleda pöördülesandena, kus sisendiks on osakeste poolt jäetud mõõteandmed kui suur punkt pilv, ja väljundiks on kõikide osakeste omadused. Niisugust ülesannet on vaja lahendada arvutuslikult efektiivselt, et oleks võimalik analüüsida suuri andmehulki, ning füüsikaliselt võimalikult täpselt, et rekonstrueeritud andmetega oleks võimalik läbi viia täppismõõtmisi.

Miks on andmete analüüsimine keeruline?

Sellise ülesande lahendamiseks kasutatakse osakestefüüsika eksperimentides rekonstrueerimisalgoritme. Tänapäeval kasutusel olevad algoritmid on koostatud üldiselt LHC algusperioodil, st rohkem kui 15 aastat tagasi, ja nii, et neid oleks arvutis lihtne kirja panna.

See tähendab, et praegused algoritmid põhinevad praktilistel reeglitel ja nn retseptidel, kuidas pörkeandmeid rekonstrueerida. Seal on tehtud olulisi lihtsustusi, et arvutus oleks tehtav. Praegused algoritmid kombineerivad osakeste poolt jäetud signaalid igas detektori kihis eraldi ning hiljem korreleerivad ühe pörke käigus registreeritud signaalid omavahel, et mitmekihilises detektoris kokku panna ühe pörke osakeste terviklik pilt.

Niisugune lähenemine töötab hästi väikese heledusega katsefaasis, st kui osakesi ei ole ühes pörkes korraga liiga palju, kuid suure osakeste arvu korral muutub see kombinatoorselt keerukaks. See tähendab, et kõiki ühes pörkes mõõdetud andmepunkte tuleb potentsiaalselt omavahel võrrelda, mille tõttu kasvab vajalike arvutuste hulk proportsionaalselt osakeste arvu ruuduga.

Samuti ei ole selliste heuristiliste meetodite täpsus alati parim, kuna keerukate ja osaliselt kattuvate või puudulike signaalide rekonstrueerimise korral ei pruugi lihtsad retseptid töötada kuigi efektiivselt. LHC kõrgeledusfaasis suureneb iga prootonpörke andmehulk kuni viis korda, kogutud andmete hulk suureneb kokku kuni kümme korda.

Lisaks paigaldatakse uued, kõrgema lahutusvõime ja rohkemate kanalitega detektorid, et füüsikaprotsesse veelgi täpsemalt mõõta. Nende granulaarsete detektorite andmete rekonstrueerimine on oluliselt keerukam, seni kasutusel olevad algoritmid tuleb selle jaoks kohandada või luua uued algoritmid.

Mis on tehisaru ja mis on sellest kasu osakestefüüsika andmete analüüsimisel?

Masinõpet ja tehisnärvivõrke on CERN-i eksperimentides rakendatud juba 1990-ndatest, kuid ligipääs kõrgjõudlusarvutusele ja suurtele andmetikele on toonud võimalike ülesannete keerukusse ja lahenduste täpsusesse arenguhüppe, mida ma oma tööde seerias ära kasutasin.

Masinõpe võimaldab piisavalt suurte andmehulkade korral keerukaid pöörd-ülesandeid lahendada uudsel moel. Selle asemel, et käsitsi luua võimalikult täpset heuristilist mudelit, kasutatakse suurte vabade parameetrite arvuga üldiseid mudeleid ehk tehisnärvivõrke, mida sobitatakse suurte andmehulkadel, kasutades optimeerimiseks arvutuslikke väakriteeriume. Teadustööde seeria sisuks on uudsete masinõppel põhinevate rekonstrueerimisalgoritmide loomine ja juurutamine CERN-i eksperimentides. Need algoritmid on varasematest täpsemad, taaskasutatavad uute katsetingimuste ja detektorite jaoks ning võimaldavad ära kasutada uudseid arvutussüsteeme, sh graafikaprotsessoreid ja superarvuteid.

Osakestefüüsika andmeanalüüsis tuleb suurte mahtude tõttu kasutada kõiki olemasolevaid arvutusressursse, sh graafikakiirendeid ja superarvuteid, kuid praeguste heuristiliste algoritmide ümberkirjutamine on töömahukas ja vajab uudset lähenemist. Samaaegselt on arenduses uued kiirendid, nagu Future Circular Collider (FCC), International Linear Collider (ILC) või müüonpõrguti Muon Collider, ja uudsed detektoritehnoloogiad, mis vajavad vastavalt innovatsiooni ja uut lähenemist andmeanalüüsis. Siin on üks võimalikke lahendusi masinõpe, mis sobib hästi superarvutitele ja võimaldab paindlike ja üldiste algoritmide arendamist.

Senimaani on masinõpet peamiselt kasutatud konkreetsete hästidefineeritud klassifikatsiooniülesannete lahendamiseks, näiteks b-kvarkidest tekkinud osakeste kaskaadide ehk jugade eristamiseks. Antud juhtudel on optimeerimisülesande sisuks kindla fikseeritud struktuuriga sisendist ehk joas olevatest osakestest jah-ei tüüpi otsuse tegemine, mis on masinõppes küllaltki lihtsalt lahenduv ülesanne. Sellist tüüpi masinõppemeetodite optimeerimiseks kasutatakse tehisandmeid, enamasti juba laialdaselt läbi uuritud jugadel põhinevaid näidisandmestikke.

Kuidas õpib tehisaru andmetest uusi mustreid ära tundma?

Tehisaru ehk masinõppe eripära seisneb selles, et käsitsi koostatud reeglite ja programmide asemel õpib arvuti matemaatilise optimeerimise abil andmetest ise sobivaid reegleid. Need võivad olla oluliselt detailsemad ja täpsemad, kui oskab kirja panna inimene. Vaja on aga andmestikke, kus masinõppe mudeleid optimeerida ehk treenida. Osakestefüüsikas kasutatakse selleks tehisandmeid, mida tekitame omakorda spetsiaalsete arvutiprogrammidega.

Nimelt on võimalik praegu teadaolevad füüsikaseadused, st osakeste tekkimise, ainega interakteerumise ja lagunemise tõenäosused kirja panna võrranditena,

mis on katseandmetega detailselt kontrollitud. Neid võrrandeid on võimalik lahendada Monte Carlo meetodil ehk arvuti abil pseudojuhuslikke arve genereerides ja seeläbi tekitada üprisri realistikke tehisandmeid, kus meil on kõik toimunud protsessid detailselt teada.

Kõik osakestefüüsika mõõtmised on sisuliselt mõõteandmete ja tehisandmete võrdlused, kus nende kahe erinevusest järeldatakse midagi pärismaailmas toimunud protsesside kohta. Praegusel ajal on tehisandmete täpsus väga suur ja seetõttu on osakestefüüsika eelisolukorras masinõppe mudelite arendamisel, kuna meil on võimalik kasutada ülimalt täpseid ja realistlikke tehisandmestikke, et masinõppe mudeleid arendada ja testida.

Osakestefüüsika andmestike eripära seisneb selles, et igas osakeste põrkes tekib juhuslik arv uusi osakesi, samuti tekib iga osakese lagunemisel või ainega interakteerumisel jällegi juhuslik arv madalama energiaga osakesi. Kui seda võrrelda näiteks pildiandmetega, kus igal pildil on fikseeritud arv pikseleid, siis osakestefüüsikas on igal järgneval mõõdetud põrkel eelnevast erinev arv tekkinud ja seega ka mõõdetud osakesi. See tähendab, et osakestefüüsikas tuleb kasutada selliseid masinõppe algoritme, mis suudavad töödelda varieeruva arvuga sisendeid ja väljundeid. Seda tüüpi algoritmide on näiteks punktipilvede või keele jaoks loodud mudelid, mida on võimalik osakestefüüsika ülesannete lahendamiseks kohendada.

Põrkes tekkinud kõrge energiaga osakesed lagunevad ja tekitavad palju madala energiaga osakesi. Seda protsessi nimetatakse hadroniseerumiseks ehk osakeste jugade tekkimiseks. See tähendab, et tekkinud algosakese omadused määratakse temast tekkinud madala energiaga osakeste vahel laiali. Kui soovime määrata algosakese omadusi, näiteks kas tekkis kolmanda põlvkonna kvark ehk b-kvark või esimese põlvkonna kvark ehk u-kvark, siis tuleb vaadata tekkinud osakeste joa komponente ja nende omavahelist jaotust. Nii on võimalik laguproduktide jaotusest statistiliselt välja lugeda, mis tüüpi algosake joa tekitas. Seda analüüsi nimetatakse jugade klassifitseerimiseks ja siin on masinõppe olnud väga kasulik tööriist, mis on oluliselt suurendanud analüüsides täpsust. Sellisel juhul on ülesandeks joa muustrist jah-ei tüüpi klassifikatsiooniülesande lahendamine, mis on masinõppe abil kergelt lahendatav probleem.

Kui soovime aga lihtsalt jugade klassifitseerimise jah-ei tüüpi ülesande lahendamisest edasi liikuda tervete põrgete täielikule rekonstrueerimisele, siis on vaja uut tüüpi meetodeid. Masinõppe peab andma mitte ainult ühe väljundi, vaid varieeruva arvu väljundeid iga rekonstrueeritava osakese kohta. Veelgi enam, iga osakese kohta on meil vaja rekonstrueerida mitmeid erinevaid mõõdetavaid: energia, mass ja teised relativistliku impulsi vektori komponendid, osakese tüüp (näiteks kas tegu oli elektroni või footoniga), osakese laeng jt. Seega tõuseb masinõppeülesande keerukus oluliselt, kuna ühe väljundi asemel tuleb õigesti rekonstrueerida kuni kümneid tuhandeid mõõdetavaid suuruseid, millel igaühel on oma iseärane jaotus.

Ülesande lahendamiseks tuli rakendada optimeerimiskriteeriumi, mille vastu masinõppe mudelit treenitakse. Nimelt on võimalik tehisandmete korral leida ideaalpilt kõikidest osakestest, mis meil tuleks antud mõõteandmete korral rekonstrueerida. Masinõppe mudel üritab vastavalt sisendist seda ideaalpilti taastoota ning optimeerimiskriteerium mõõdab, kui kaugel on konkreetne rekonstrueeritud pörge ideaalpildist.

Miks anti 2024. aasta Nobeli preemiad füüsikas ja keemias tehisaru avastuste eest?

2024. aasta Nobeli preemia füüsikas anti tehisnärvivõrkudega seotud avastuste eest biofüüsikule John Hopfieldile ja arvutiteadlasele Geoffrey Hintonile. Mõlema teadlase töö kasutas teadmisi füüsikast, et arendada välja täiesti uut tüüpi andmetöötluse algoritm ehk tehisnärvivõrk.

Hopfield kasutas statistilise füüsika meetodeid, et luua algoritm, mis suutis eelnevalt nähtud andmeid meelde jätta, kodeerides need enda sisestruktuuri ja siis hiljem neid andmeid näidete varal uuesti meelde tuletada. Statistiline füüsika tegeleb paljudest väikestest osadest koosnevate süsteemide kirjeldamisega. Selline statistiline süsteem on näiteks väikestest magnetitest koostatud võre, mille iga element mõjutab oma naabreid. Hopfield näitas, et kui sisendandmed kodeeritakse võresõlmedevahelistesse seostesse ja nendest lähtuvasse võre energiatasemetesse, siis mõnevõrra kahjustunud sisendandmete korral suudab võre siiski taastoota esialgse mustri, kuna võres tekkinud energiatasemed suunavad võre vigaste või osaliste andmete korral esialgsesse olekusse ehk treeningandmetest tekitatud mustrisse tagasi. Niimoodi oli võimalik luua statistilise füüsika meetodite abil mälu omadustega algoritm, mis suutis salvestada keerukaid andmeid võre ehk tehisnärvivõrgu struktuuri.

Hinton omakorda täiendas seda lähenemist juhule, kus võre elementidel on pidevad väärtused. Seda süsteemi kirjeldab Boltzmanni jaotus, mida kasutatakse ka statistilises füüsikas gaasi omaduste kirjeldamiseks. Hintoni arendatud närvivõrk ehk Boltzmanni masin suutis ära tunda samast objektist uusi, seni nägemata pilte, mis näitas, et tehisnärvivõrgud suudavad lahendada keerukaid pärismaailma ülesandeid, nagu objektituvastus. Edasi arendas ta välja meetodi, kus tehisnärvivõrgule lisati mitmeid peidetud kihte, mis võimaldasid varasemast keerukamat andmetöötlust. Samuti arendas ta meetodi, kus närvivõrgu omadusi optimeeriti matemaatiliselt efektiivse gradientlaskumise meetodi abil, mida oli vaja, et mitme kihiga ehk paljude vabade parameetritega närvivõrke pärismaailma ülesannetega treenida. Füüsikateadmised võimaldasid seega Hopfieldil, Hintonil ja teistel välja arendada täiesti uusi andmeanalüüsi algoritme.

Läbimurre paljude kihtidega närvivõrkude kasutamises toimus aastal 2012, kui AlexNeti-nimeline tehisnärvivõrk ületas esmakordselt suure, pärismaailma piltide klassifitseerimise ülesande lahendamisel ImageNeti senised tulemused. Kuna sellel tehisnärvivõrgul oli umbes 60 miljonit vaba parameetrit, siis tuli

modeli treenimiseks kasutada uudseid arvutusmeetodeid, nimelt graafikakaarte, mis võimaldavad kiiresti ja paralleelselt läbi viia väga suure arvu operatsioone ja seeläbi selliseid massiivseid tehisnärvivõrke suurtel andmetel treenida.

Praeguseks oleme jõudnud meetodite arenguga sinnamaale, et närvivõrke võib kasutada uue teadmise loomiseks. Hea näide on siin 2024 välja antud Nobeli preemia keemias, mis anti uudse, närvivõrkudel ja suurandmetel põhineva valkude struktuuri määramise algoritmi AlphaFold loojatele John Jumperile ja Demis Hassabisele ning arvutusmeetodite abil uute valkude loojale David Bakerile.

Valkude struktuuri määramine on arvutuslikult keerukas ülesanne, kus tuleb leida valkude ehituskivide ehk aminohapete jadast valgu kolmemõõtmeline ehitus. Suurtest andmehulkadest välja kaevatud närvivõrgul põhinev mudel osutus siin oluliselt täpsemaks senimaani käsitsi arendatud mudelitest, kuna teisendusnärvivõrkude ehk *transformer*-tüüpi algoritmi kasutamine võimaldas algoritmil kõikide seni tuntud valkude struktuuri andmebaasis leida seaduspärad, mida varem ei osatud näha.

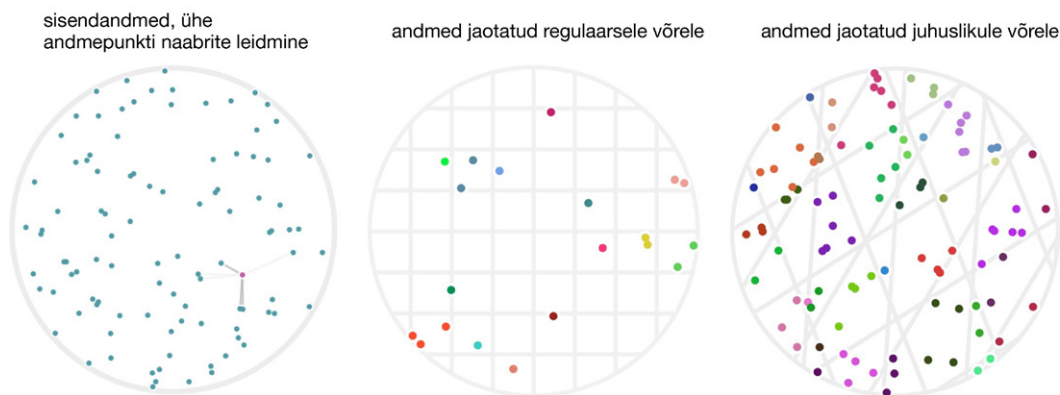
Siit ilmneb muster teaduses: ühe valdkonna väikesed avastused võivad tuua suuri avastusi teistes, esmapilgul mitteseotud valdkondades. Mõnikord tuleb edasiminekuks siis, kui olemasolevale probleemile lähenetakse uuel moel.

Mis on graafidel põhinevad ja *transformer*-tüüpi võrgud?

Kõikide pörkes olevate osakeste üheaegseks rekonstrueerimiseks on tarvis luua uus kontseptuaalne lähenemine, kus on võimalik tehisandmetes seostada pörkes tekkinud stabiilsed osakesed, nende poolt jäetud signaalid ja nende omavahelised seosed ühtses, ilma ette fikseeritud suurusega andmestruktuuris.

Selliste määramata arvuga, kuid omavaheliste seostega pörkeandmete töötlemiseks masinõppe abil oli vaja rakendada uudseid meetodeid, sh graafidel põhinevaid närvivõrke ja teisendusnärvivõrke (*transformer*), mida kasutavad ka tänapäevased keelemudelid, nagu ChatGPT või eelmainitud valkude voltija AlphaFold. Peamiseks uuenduseks minu töös oli senimaani keele, molekulide ja osakeste jugade töötlemiseks kasutatud tehisnärvivõrkude edukas kohandamine tervete pörkesündmuste üheaegseks töötlemiseks.

Kuna ühe töödeldava andmepunkti ehk pörke detailsus kasvas sellega seoses kuni 50 korda, oli vaja rakendada uudseid arvutusmeetodeid graafide konstrueerimisel ja analüüsimisel, et arvutusi praktikas suures mahus läbi viia saaks. Siin raskendasime Google'i teadlaste poolt välja töötatud lokaaltundlikku räsimit, mis oli antud meetodi üks esmakordseid rakendamisi osakestefüüsikas. Lokaaltundlik räsimine tähendab, et suur arv sisendandmepunkte pakitakse vastavalt nende omavahelisele matemaatilisele kaugusele või lähedusele mitmemõõtmelise võre peal asuvasse kastidesse.



Allikas: https://unboxresearch.com/articles/lsh_post1.html

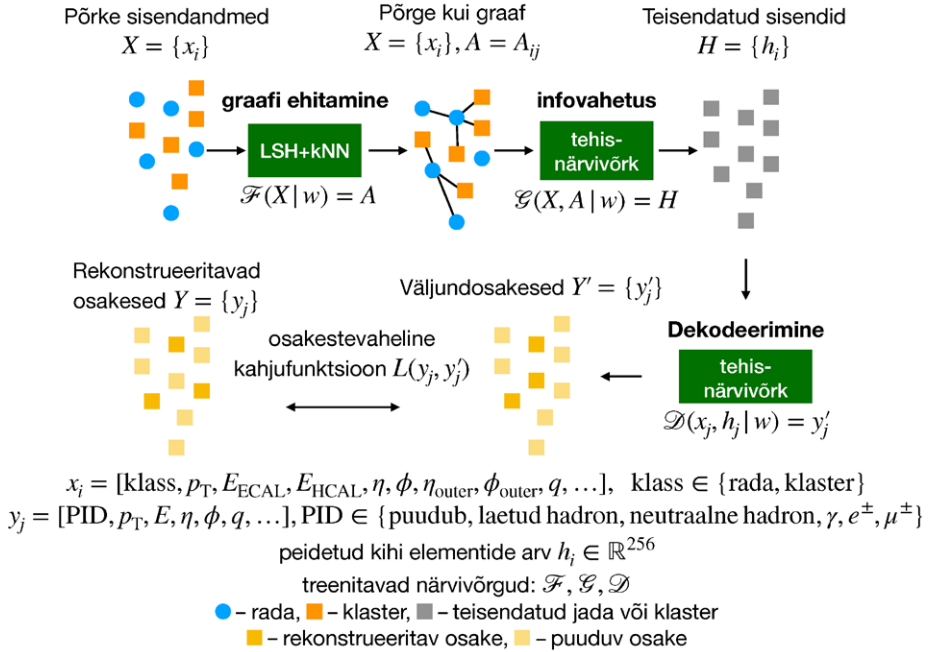
Joonis 2. Lokaaltundliku räsamise abil suurandmetel efektiivsete arvutuste läbiviimise skeem. Andmepunktid jagatakse räsifunktsiooni alusel lähedaste punktidega samasse võrekasti. Paaridevahelised arvutuslikult keerukad tehted viiakse läbi ainult ühes kastis olevate andmepunktide vahel. Väga suurte sisendmõõtmete korral on võimalik kasutada juhuslikku võret, mis annab andmepunktide jaotamisel hea tulemuse.

Arvutuslikult keerulised tehted tehakse nüüd ainult nende andmepunktide vahel, mis on lähedastikku ehk langevad ühte kasti, samas kui ilma selle meetodita peaks arvutuslikult keerukaid tehteid tegema üle kõigi andmepunktide paaride, mis võib olla arvutuslikult väga kallis. Nüüd on meetod leidnud osakestefüüsikas laiemat rakendust selliste arvutuslikult keerukate ülesannete lahendamiseks. Seega olime ühed esimestest, kes näitasid, kuidas osakestefüüsika keerukaid ja väga suuri sisendandmeid analüüsida masinõppega arvutuslikult efektiivselt.

Kuidas arendatakse Eestis närvivõrkudel põhinevat andmetööstlust?

Aastatel 2020–2024 on minu loodud ja juhitud rahvusvaheline (CERN, Caltech, UCSD jt) meeskond välja töötanud ja juurutanud uudse, masinõppel põhineva lähenemise osakeste rekonstrueerimiseks detektorites ja seeläbi loonud Eestis uue uurimissuuna, mis on rahvusvahelisel tasemel edukas. Oleme loonud uue analüüsimeetodi, mis asendab suurandmete analüüsis käsitsi kirjutatud heurisilised algoritmid hästidefineeritud, efektiivsete ja testitud tehisinärvivõrkudega.

Järgnevak스 uuenduseks on närvivõrgu väljundi komplekskuse tõstmine pörgete rekonstrueerimiseks vajalikule tasemele. Senimaani saadi närvivõrgust tüüpiliselt ühe sisendandmepunkti, näiteks ühe osakeste joa kohta üks kvantitatiivne väljund, kuid pörke rekonstrueerimiseks oli vajalik töödelda korraga kogu pörkes registreeritud signaalide kogum ning väljastada korraga kõikide rekonstrueeritud osakeste omadused. Siin lõime aluse oluliselt keerukamate osakestefüüsika ülesannete lahendamiseks masinõppe abil.



Joonis 3. Meie poolt välja arendatud graafidel põhinevat närvivõrku kasutatav osakeste rekonstrueerimise skeem. Ühes põrkes rekonstrueeritud signaalid moodustavad sisendandmed. Andmete vahel koostatakse lokaaltundliku räsamise ja lähimate naabrite algoritmi abil graaf, mida kasutatakse graafidel põhineva tehisnärvivõrgu abil andmepunktide vahel info vahetamiseks. Niimoodi kodeeritakse andmepunktidest nende naabrite informatsioon. Edasi dekodeerib järgmine tehisnärvivõrk andmed rekonstrueeritud osakesteks, mida võrreldakse soovitud osakestega tehisandmetes, kasutades osakestepõhist kahjufunktsiooni. Kõiki tehisnärvivõrke treenitakse üheaegselt miljonitel põrgetel.

Rekonstrueeritud ja tõeliste osakeste arvutuslikuks võrdlemiseks oli vaja luua optimeerimiskriteerium ehk nn kahjufunktsioon, mida saaks kasutada närvivõrkude treenimisel sellel uudse ülesande puhul. Tööde seerias lõime ja juurutasime uut tüüpi kriteeriumi, mis põhineb füüsikalisel ideel, kus igale rekonstrueeritavale osakesele seatakse tehisandmete alusel vastavusse peamine sisendsignaal. Uudne meetod võimaldab optimeerida osakeste rekonstrueerimise närvivõrgu väljundit väga suurtes andmestikes praktiliselt ja kiirelt.

Miks on tehisarul põhinevad meetodid paremad olemasolevatest algoritmidest?

Tehisnärvivõrkudel põhineva algoritmi eeliseks on suur paindlikkus, st sama algoritmi on võimalik väikeste muudatustega taaskasutada uute detektorite jaoks. Tavapäraseid, käsitsi programmeeritud ehk reeglitel põhinevaid algoritme on

keerukas kohandada, kui detektori ehitus või struktuur peaks muutuma, kuid närvivõrke saab kergesti uutel andmestikel uuesti optimeerida ehk treenida. See võimaldab juba praegu välja arendatud metoodikat kasutada tulevikus planeeritud detektorite arendamiseks ja annab võimaluse panustada LHC asenduseks välja pakutud tulevikupõrguti FCC, lineaarkiirendite või ka müüonpõrguti arendusse.

Põhiline uudne tulemus tööde seerias on uudse masinõppeülesande püstitamises ja edukas lahendamises, kus varasem reeglitel põhinev rekonstrueerimisalgoritm asendatakse masinõpitud algoritmiga, mis on täpsem, arvutuslikult efektiivsem ja paindlikum.

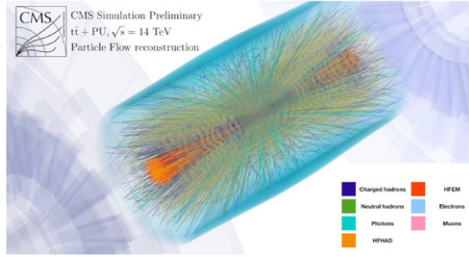
Aastal 2021 avaldatud artiklis (3) saavutas minu juhitud meeskond esmased tulemused osakeste rekonstrueerimises masinõppega uudsel tehisandmestikul. Masinõppemudelil olid olemas põhilised olulised ja uudsed komponendid: terve hadronpõrke sündmuste osakeste rekonstrueerimine varieeruva osakeste arvu korral, arvutuslik skaleeruvus suurtele andmestikele ja heuristilisest algoritmist suurem täpsus. Tulemuse olulisus seisnes seda tüüpi mudeli esmakordses rakendatavuse tõestamises.

Samaaegselt testib ja arendab meie meeskond uut metoodikat CERN-i kompaktse müüonsolenoidi CMS (Compac Muon Solenoid) detektoris, kus on võimalik antud lähenemist katsetada päriselu tingimustes, sh CMS-i katseandmetes. Tulemused on avaldatud rahvusvahelisel konverentsil ACAT 2021 ja 2022 (4–5). Selle tööga seoses olin aastatel 2020–2022 CMS-i rekonstrueerimistarkvara meeskonna koordinaator. See tähendas igapäevast tööd kogu detektori osade tarkvaraarenduse koordineerimisel, et tükid sobiksid omavahel loogiliseks tervikuks ning et detektor oleks võimeline aastal 2022 taaskäivitatud LHC kiirendi poolt tekitatud põrked piisava täpsuse ja efektiivsusega üles võtta. Seoses andmetöötluse optimeerimisega saavutatud tulemustega tunnustati minu panust ka CMS-i kollaboratsiooni (CMS Collaboration – CERN-is CMS-i eksperimenti läbi viiv konsortsium) auhinnaga.³⁶

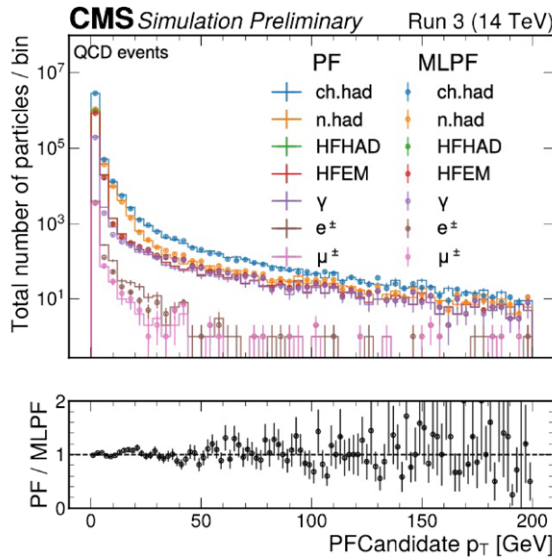
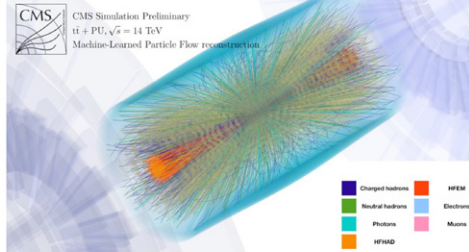
Tavapäraselt arendatakse ja testitakse osakestefüüsikas selliseid masinõppe meetodeid ainult lihtsustatud tehisandmetel. Meie meeskond arvab, et meetodit tuleb rakendada pärismaailmas võimalikult realistlikel andmetel, sest ainult nii saab veenduda, et tegemist on ka tegelikult töötava lahendusega. Oleme näidanud, et metoodikat saab edukalt integreerida CMS-i katseseadmesse ja see annab olemasolevast meetodist täpsemaid tulemusi. Esialgsed tulemused on ette kantud rahvusvahelistel konverentsidel, mille eelduseks oli CMS kollaboratsiooni sisese retsensiooni läbimine. Praegu on ettevalmistamisel CMS kollaboratsiooni artikkel, mille valmimisel on tegu esmakordse ja uudse tulemusega osakestefüüsika andmeanalüüsis.

³⁶ <https://cms.cern/news/cms-award-2021>

tava-algortm: PF



masinõpitud algortm: MLPF



Joonis 4. CERN-i CMS-i detektoris saavutatud esialgsed tulemused masinõpitud osakeste rekonstrueerimise algoritmi MLPF jaoks võrdluses tavaalgoritmiga PF. Tulemused on samaväärsed, mis tähendab, et meie poolt arendatud algoritmi on võimalik päriselu tingimustes rakendada.

Joonise terminite tõlked: CMS simulation preliminary – CMS-i simulatsioon, esmased tulemused; Particle Flow reconstruction – osakestevoo algortm ehk tava-algortm; Machine-Learned Particle Flow reconstruction – masinõpitud osakestevoo algortm; QCD events – QCD pörked; Total number of particles / bin – osakeste arv; charged hadron ehk ch.had – laetud hadron; neutral hadron ehk n.had – neutraalne hadron; photon – footon; electron – elektron; muon – müüon.

Järgnevas, aastal 2024 avaldatud artiklis (2) näitasime, et meetodi täpsust ja tõepärasust saab oluliselt suurendada, kuna kasutasime esimest korda tulevikudetektor CLIC jaoks välja arendatud detailset kiirguse-materjali simulatsioonitarkvara. Näitasime, et meie loodud käsitusviisi saab kasutada paljude erinevate praeguste ja tulevikudetektorite põrgete täpselt ja paindlikult rekonstrueerimiseks. See on väga oluline, kuna tuleviku tarbeks planeeritud detektorite arendusel on vaja kiiresti testida, kas konkreetne detektoritehnoloogia võimaldab simulatsioonides paremaid rekonstrueerimise tulemusi saada. Käsitsi kirjutatud tarkvara optimeerimine uutes tingimustes on aga äärmiselt aeganõudev. Samas on võimalik masinõppel põhinevaid meetodeid kiiresti uute andmete peal treenida ja saada nii peaaegu optimaalsed tulemused.

Millised on Eesti digitiigri võimalused CERN-i täisliikmena?

Arvestades 2024. aasta Nobeli preemiat füüsikas (tehninärvivõrkude alusmeetodite arendamine) ja keemias (masinõppe rakendamine valgustruktuuride määramisel), on masinõppe rakendamine osakestefüüsika andmete analüüsis äärmiselt ajakohane. Võttes arvesse Eesti liitumist CERN-iga ning Eesti kui digiriigi strateegilist arengusuunda, on masinõppel ja tehisintellektil põhinevate meetodite arendamine CERN-i teadusrakendusteks Eestile oluline võimalus rahvusvahelisel teadusmaastikul.

Oleme loonud Eestis uue, rahvusvahelisel tasemel oskusteabe ja ka tööühma masinõppel põhinevate osakestefüüsika rekonstrueerimisalgoritmide arendamiseks. Oleme laiendanud tegevust tau leptonite rekonstrueerimisalgoritmide arendamiseks (1, 6). Tau leptonid on kui hästi massiivsed elektronid, mis lagunevad kiiresti spetsiifilise mustriga osakeste jugadeks. Nad on eriti olulised Higgsi bosoni eneseinteraktsioonide uurimisel, kuna Higgsi boson laguneb küllaltki arvestataval määral massiivseteks tau leptoniteks, samas on need osakesed nii harvad, et neid on võimalik muudest tekkinud osakestest üsna kergesti eristada.

Varem rekonstrueeriti ka tau leptoneid peamiselt käsitsi kirjutatud reeglipõhiste algoritmidega. Need algoritmid ei ole tingimata väga kiired ega ka väga täpsed. Meie välja arendatud meetod võtab sisendiks otse joas olevad osakesed ja nende mõõdetud omadused ning lahendab korraga ära mitmetahulise masinõppeülesande. Nimelt määrame *transformer*-tüüpi mudeliga korraga ära, kas antud juga pärines tau leptonist või mitte (jah-ei tüüpi klassifikatsiooniülesanne), mis tüüpi tau leptoni lagunemisega oli tegu (mitmikklassifikatsiooni ülesanne) ning mis oli joa tekitanud tau leptoni esialgne energia (regressioon- ehk lähendusülesanne).

Seega on minu juhitud rahvusvahelise meeskonna uudne lähenemine ja edukas teostus avanud uue arengusuuna osakestefüüsika andmete rekonstrueerimises ning masinõppel põhineva uudse teadusarvutuse metoodikas. Metoodika on eriti oluline seoses CERN-i tulevaste uuendustega ja annab Eesti teadlastele võimaluse panustada CERN-i teadustegevusse rahvusvahelisel tasemel. Metoodikat saab edasi arendada, et luua üldine osakestefüüsika põrgete rekonstrueerimise

alusmudel, mida saab taaskasutada paljude ülesannete jaoks. Arvestades masinõppe ja tehisintellekti kiiret arengut, on seda tüüpi ülesannete püstitamise ja lahendamise oskus oluline ka pädeva tööjõu koolitamisel tehisintellekti alal.

Millised võimalused on Eesti noortel CERN-i tegevustes osaleda?

CERN on Euroopas ja maailmas osakestefüüsika eesliiniteaduse lipulaev. Seal töötavad oma valdkonna tipud, infovahetus on vaba ja vahetu. Seega on CERN suurepärane koht, kus oma oskusi proovile panna ja ennast arendada nii füüsika, arvutiteaduse kui ka üldiselt suurte projektide ja rahvusvahelise koostöö toimimise vallas.

Eriti saab rõhutada masinõppe praegust ja järjest kasvavat olulisust CERN-i tegevustes nii eesliiniteaduses kui ka üldiselt protsesside optimeerimisel. Tehisintellekti meetodite arendamiseks on vaja väga suuri andmehulkasid ja oskust nendega ümber käia. Üldiselt on sellised andmestikud ainult suurettevõtetel, nagu Meta, Google, Microsoft jne, kes on kogunud need oma kasutajatelt ega anna nendele ligipääsu just igaühele.

Samas on CERN-il maailma suurimad avaandmestikud, mis on vabalt kasutada kõikidele soovijatele koos väga detailsete metaandmetega. Kuna need andmed on kogutud füüsikaeksperimentidest, siis ei kaasne nendega selliseid privaatsusega seotud eetilisi küsimusi, mis paratamatult on olulised isiklike andmete puhul. Samuti on füüsikaseaduste näol olemas täpne arvutuslik mudel, mis oskab neid andmeid küllaltki täpselt seletada. Me suudame teadaoleva füüsika alusel luua ka väga täpseid tehisandmeid, mida saab kasutada tehisaru mudelite treenimiseks ja testimiseks. Seega on CERN suurepärane kasvulava tehisaru tehnoloogiale ja teadmisele, kus osavõtt ei ole kuidagi piiratud.

Milliseid teadmisi oleks vaja, et CERN-i andmeanalüüsis ja masinõppe arenduses panustada?

Kõige olulisem on oskus loogiliselt mõelda ja oma ideid praktiliselt rakendada. Tänapäeval tähendab see programmeerimise ja andmete analüüsimise oskust.

Siin tuleb juba praegu appi tehisaru, mis oskab üpriski hästi tõlkida inimkeelsed juhised masinkeelde, näiteks laialdaselt kasutatavasse Pythoni programmeerimiskeelde. Küll aga ei tea tehisaru, mis ülesannet on vaja lahendada, samuti ei oska ta sõnastada piisava täpsusega soovitud küsimusi ega analüüsida võimalikke vigu või tekkivaid probleeme. Seega soovitatakse harjutada loogilist mõtlemist ning küsimuste täpset ja detailset sõnastamist, samuti praktilist programmeerimist koos tehisaruga konkreetsete probleemide lahendamise jaoks. Lisaks on vaja harjutada oskust, kuidas saada aru, kas tehisaru ajab sulle pada nii, et suu suitseb.

Oluline on teada, kuidas püstitada küsimusi ja leida neile andmetest vastuseid. Enamasti on kõige keerulisem küsida õige küsimus, millele vastuse leidmine nihutab teadmiste piiri veidi edasi. Piisavalt täpses küsimuses on tihti juba ka vastuse leidmise alge, aga see vajab ka täpset mõtlemist.

Kuna Eestis on käivitumas tehisintellekti ehk TI-hüpe, siis on siin väga hea võimalus ära kasutada ka Eesti ligipääsu CERN-ile ja sealsetele eesliiniarendustele rakendusliku tehisaru valdkonnas. Eesti teadlased ja tudengid saavad osaleda projektides, samuti kandideerida CERN-i töökohtadele IT, andmeteaduse ja tehisintellekti valdkondades. See on võimalus koolitada uue põlvkonna tehisarutsinere ja rakendada neid omakorda kõrgtehnoloogilise tööstuse arendamises.

VIITED

1. Lange, T., Nandan, S., Pata, J., Tani, L., Veelken, C. Tau lepton identification and reconstruction: a new frontier for jet-tagging ML algorithms. *Computer Physics Communications* 298 (2024), doi: 10.1016/j.cpc.2024.109095
2. Pata, J., Duarte, J., Mokhtar, F., Wulff, E., Yoo, J., Vlimant, J-R., Pierini, M., Girone, M. CMS kollaboratsiooni nimel. Machine Learning for Particle Flow Reconstruction at CMS. – 20th International Workshop on Advanced Computing and Analysis Techniques in Physics Research, ACAT 2021. *J. Phys. Conf. Ser.* 2438 012100, doi: 10.1088/1742-6596/2438/1/012100
3. Pata, J., Duarte, J., Vlimant, J-R., Pierini, M., Spiropulu, M. MLPF: efficient machine-learned particle-flow reconstruction using graph neural networks. *Eur. Phys. J. C* 81, 381 (2021), doi: 10.1140/epjc/s10052-021-09158-w
4. Pata, J., Mokhtar, F., Duarte, J., Wulff, E., Pierini, M., Vlimant, J-R. CMS kollaboratsiooni nimel. Progress towards an improved particle flow algorithm at CMS with machine learning. – 21st International Workshop on Advanced Computing and Analysis Techniques in Physics Research, ACAT 2022. CMS Detector Performance Note 2022, URL: <http://cds.cern.ch/record/2842375>
5. Pata, J., Wulff, E., Mokhtar, F., Southwick, D., Zhang, M., Girone, M., Duarte, J. Improved particle-flow event reconstruction with scalable neural networks for current and future particle detectors. *Nature Commun Phys* 7, 124 (2024), doi: 10.1038/s42005-024-01599-5
6. Tani, L., Seeba, N-N., Vanaveski, H., Pata, J., Lange, T. A unified machine learning approach for reconstructing hadronically decaying tau leptons. *Computer Physics Communications* 307 (2024), doi: 10.1016/j.cpc.2024.109399



Foto: Arno Mikkor

Joosep Pata

Sündinud 24. mail 1990 Tartus

2008 Miina Härma Gümnaasium

2012 Tartu Ülikool, füüsika (BSc)

2014 Tartu Ülikool, füüsika (MSc, *cum laude*)

2018 ETH Zürich, füüsika (PhD, ETH medal)

Pärast doktoriõpinguid töötas aastatel 2018–2020 USA-s Caltechis järel doktorina. Aastast 2021 töötab Keemilise ja Bioloogilise Füüsika Instituudis kõrge energia ja arvutusfüüsika laboris vanemteadurina. Aastatel 2020–2022 oli CERN-i CMS-i kollaboratsiooni rekonstruktsioonitarkvara koordinaator.