
TOOL WEAR PREDICTION WITH DEEP LEARNING MODELS

Gorka Dabó Aizpurua

Univ: University of the Basque Country
City: San Sebastián
gdabo001@ikasle.ehu.eus

Mikel Osoz Rivas

Univ: University of the Basque Country
City: Irún
moscoz002@ikasle.ehu.eus

1 Proiektuaren deskribapena

Proiektu honek mekanizazio prozesuetan erreminten higadura aurreikusteko deep learning estrategia ikertzen du. Erreminta bat higatuta dagoen edo fabrikatutako piezen kalitatean eta ekoizpenaren eraginkortasunean eragin esanguratsurik izan daiteke, horregatik makina bat higatuta degoen ala ez aurreikusteko gaitasuna oso erabilgarria da.

Arazo horri aurre egiteko, CNC Mill Tool Wear datu-multzoa erabiltzen da, 18 mekanizazio prozesuri buruzko informazioa duena. Helburua neurona-sareen arkitekturak inplementatzea eta ebaluatzea da, tresna higatuta dagoen ala ez zehazteko. Datuen kopuru mugatua dela eta, "utzi bat kanpo" (leave-one-out) baliozkotze gurutzatua eta "sliding-window" teknikak erabiliko dira ereduaren ebaluazioa neurtzeko.

Proiektuaren barruan sartzen dira, besteak beste, datuak egituratzeko ikuspegi egokia diseinatzea, deep learning arkitektura desberdinak aldatzea eta ebaluazio metodologia sendoa ezartzea. Emaitzak eta aurkikuntzak Colab koaderno batean inplementatuko dira, garatutako ereduaren funtzionamendua frogatzeko.

2 Arazoa

Mekanizazio industrialeko prozesuetan, erreminten higadura faktore kritikoa da, eta fabrikatutako piezen kalitatean eta enpresaren produktibitatean eragiten dezake. Higatutako tresna bat behar denengan detektatzen ez bada, akatsak sor ditzake produktuetan, atzerapenak produkzioan eta eragiketa kostuak handitu dezake.

Gaur egun, erreminten higaduraren identifikazioa eskuzko edo aurrez definitutako arauetan oinarritutako prozesu bat da, eta horrek detekzioaren zehaztasuna eta azkartasuna mugatzen ditu, bereziki automatizazio eta eraginkortasun maila handiak eskatzen dituzten industria eredu modernoetan, 4.0 Industria bezala. Errebisio periodikoak egiten dituzten enpresa batean, makinaren higadura ez dute antzematzeko teknika datorren hurrengora arte eta hain konpetitiboa den sektore batean honek galera handiak sor ditzake.

Proiektu honek lantzen duen arazoa deep learning-ean oinarritutako eredu bat garatzea da, tresna bat higatuta dagoen ala ez automatikoki aurreikustea ahalbidetuko duena, mekanizazio prozesuan erregistratutako parametroak erabiliz. Arazo hori sailkapen bitarreko ataza bat da, eta hainbat erronkari egin behar zaie aurre: datu erabilgarrien kopuru mugatuari edo ebaluazioaren fidagarritasunari beste batzuen artean.

3 Datuen deskribapena

Proiektu honetan erabilitako datuak Michigan-eko Unibertsitateko CNC mekanizazio ingurune batean egindako 18 esperimentuetatik datoz. Esperimentuak egiteko, argizarizko blokeak erabili ziren oinarritzko material gisa, eta CNC fresatzeko makina bat, mekanizazio-prozesuaren parametroak erregistratzeko sentsoreekin ekipatuta. Helburua erreminten baldintzak aztertzea abiaduraren eta presioaren konfigurazio desberdinetan izan zen.

Datuak bi zatitan banatzen dira, alde batetik, esperimentu guztien datu orokorrak ditugu eta bestetik esperimentu bakoitzaren datu zehatzak.

3.1 Datu orokorrak (train.csv)

Fitxategi honek 18 esperimentuetako informazio laburtua dauka, datu guztien artean informazio garrantzitsuenak hau da:

- Esperimentuaren zenbakia
- Erabilitako materiala (cera)
- Abiadura (feed rate)
- Presioa (clamp pressure)
- Erremintaren egoera (tool condition), "unworn" (ez-higatua) edo "worn" (higatua)
- Tresnaren ikuskarpen bisualaren emaitza

3.2 Esperimentuan datuak (experimentxx.csv)

Fitxategi bakoitzak 100 ms-ko tarteetan erregistratutako neurketak ditu mekanizazio-prozesuetan. Neurketa horietan sartzen dira:

- Posizioa X, Y eta Z. ardatzetan
- Torlojuaren biraketa (spindle)
- Ueko abiadura (current feed rate)
- Exekututzen ari den programa zenbakia

4 Diseinua eta inplementazioa

4.1 Sare arkitektura

Proiektu honetan inplementatutako sare neuronalak, HybridCNNRNN izenekoak, sare konboluzionalen (CNN) eta sare errekurrenten (RNN) ezaugarriak konbinatzen ditu, mekanizazioaren prozesuan datuetan dauden ezaugarri lokalak eta mendekotasun sekuentzialak modu optimo batean aprobeztatzeko.

4.1.1 Geruza Konboluzionalak

Geruza konboluzionalak ezaugarri lokal esanguratsuenak ateratzeko diseinatu daude. Lehen geruza konboluzionalan sarrera-ezaugarrien kopurua 47 da, 32 iragazki ditu eta padding 2 tamaina ertaineko tokiko patroiak harrapatzeko. Bigarren geruza konboluzionalan 64ra igotzen da iragazkien kopurua, ezaugarri abstraktuagoak ateratzeko. Iragazki txikiagoa erabiltzen du (kernel size = 3), neurriak gordetzen dituen padding batekin. Geruza konboluzional bakoitzaren ondoren batch normalizazioa dago entrenamendua egonkortzeko, ReLU aktibazio bat, eta maxpooling eragiketa bat dimentsionaltasuna murrizteko. Azkenik, Dropout geruza bat aplikatzen da 0.3 probabilitatearekin overfitting saihesteko.

4.1.2 Geruza errekurrentek (LSTM)

Bi norazko LSTM sareak datuetan epe luzerako erlazioak sortzeko balio du. Sarreran CNNek prozesatutako ezaugarriak daude. Gero, bi LSTM geruza daude, noranzko bakoitzeko 128 unitate ezkuturrekin (256 guztira). Dropout geruza errekurrenten artean ere erabiltzen da.

4.1.3 Fully Connected geruzak

Azken geruzek ateratako ezaugarriak iragarpen bitar batean bihurtzen dituzte. Lehen geruzak dimentsionaltasuna 256tik 64ra murrizten du, ondoren ReLU aktibazioa eta dropouta. Bigarren geruzak irteera eskalarra sortzen du, eta, ondoren, Sigmoid aktibazio-funtzioa, tresna higatuta dagoen ala ez adierazten duen probabilitatea sortzeko.

Laburtuz, Sarrerako datuak berrantolatu egiten dira geruza konboluzionalek prozesatzeko. CNNek ateratako ezaugarri lokalak LSTM geruzetara pasatzen dira, denbora-erlazio zabalagoak atzemateko. LSTMen azken irteera fully connected geruzen bidez prozesatzen da, iragarpen bitarra sortzeko.

4.2 Ereduaren entrenamendua

Entrenamendu prozesua train loop funtzioaren bidez implementatzen da. Funtzio honek sare neuronalaren parametroak optimizatzen ditu eta beraien garrantzia dev multzo batean ebaluatzen du. Funtzio horren optimizaziorako, ebaluaziorako eta datuak kudeatzeko teknika aurreratuak konbinatzen ditugu, eredua sailkapen bitarreko zereginera egokitzen dela ziurtatzeko.

Funtzioak sarrera hauek hartzen ditu: entrenatu beharreko eredua, entrenamendu eta balidazio multzoetarako data loaderrak, gailua (CPU edo GPU), epoka kopurua, eta balidazio gurutzatua erabiltzeko aukerako adierazlea (leave-one-out cross-validation). Optimizazioa Adam algoritmoa erabiliz egiten da, 5×10^{-4} -ko learning rate batekin eta $L2$ erregularizazioa overfitting-a ekiditzeko.

Entrenamendu epoka bakoitzean, modeloa entrenamendu moduan (train) konfiguratu da, eta batchez batch prozesatzen da sarrera datuen multzoa. Batch bakoitza hautatutako gailura eramaten da, eta ere duaren iragarpenak kalkulatzeko erabiltzen da. Ondoren, errorea ebaluatzen da entropia gurutzatu bitarraren galera funtzioa erabiliz (Binary Cross-Entropy Loss); funtzio hori egokia delako sailkapen bitarreko arazoetarako. Sarearen parametroak Adam optimizatzailearen bidez eguneratzen dira. Prozesu horretan zehar, entrenamendu-multzoaren galera metatzen da, eta metrikak kalkulatzeko iragarpenak sortzen dira.

Entrenamendu-multzoaren iterazioa amaitu ondoren, eredua dev multzoan ebaluatzen da, eta ebaluazio-moduan konfiguratu da (eval). Etapa honetan, gradienteen kalkulua desaktibatzen da kostu konputazionala murrizteko. Balidazio multzoko iragarpenak benetako etiketekin alderatzen dira F1-score metrika kalkulatzeko, bai entrenamendu multzorako, bai balidazio multzorako.

F1-score onena lortzen duen eredua, balidazio multzoan, eredu onena bezala gordetzen da. Iterazioak amaitzean, funtzioak eredu optimizatu hori itzultzen du, iragarpen berrietan erabiltzeko prest dagoena.

4.3 Ereduaren ebaluazioa

Ereduaren ebaluazioa evaluate model funtzioaren bidez egiten da. Funtzio honi esker, entrenatutako ere duaren eraginkortasuna aztertu daiteke, leiho mailan zein esperimendu mailan. Esan beharra dago, test garaian ezin denez "sliding-window" erabili, esperimenduak leihoen tamainako azpisekuentzietan banatu ditugula entrenamendu garaian erabilitako sekuentzia luzer antzekoak erabiltzearen. Ikuspegi honek atazaren ezaugarriak hartzen ditu kontuan, non esperimendu bakoitza datuen leiho ugariz osatuta dagoen. Funtzioak bozketa (majority voting) bat erabiltzen du azken iragarpenak deribatzeke, leiho indibidualetarako egindako iragarpenetan oinarrituta, esperimendu bakoitza klasifikatzeko.

Prozedura hasterakoan, eredua ebaluazio moduan konfiguratu da, eta horrek gradienteen kalkulua desaktibatzen du karga konputazionala murrizten. Jarraian, ere duaren iragarpenak batchez batch sortzen dira, proba guztiak kontuan hartuta. Batch bakoitzerako, sarrerako leihoak hautatutako gailura mugitzen dira, eta ere duak leiho bakoitzari lotutako probabilitateak kalkulatu ditu. Probabilitate horiek iragarpen bitar batera bihurtzen dira, 0.5eko atalasea erabiliz. Iragarpenak eta benetako etiketak zerrendatan biltegitratzen dira, ondoren leiho-mailan metrikak errazago kalkulatzeko.

Leiho zein esperimendu mailako iragarpenak sortu ondoren, metrika estandarrak kalkulatu dira, hala nola zehaztasuna edo F1-score. Metrika horiek bi mailetarako kalkulatu dira (leiho eta esperimendu), eta ere duak test multzoan duen errendimenduaren ikuspegi zehatza ematen dute. Azkenik, funtzioak hiztegi bat itzultzen du, kalkulatuak metrikak, leiho eta esperimendu mailako iragarpenak eta dagozkien benetako etiketak barne hartzen dituen.

4.4 Leave One Out Cross Validation Bariantea

Gure lehenengo inplementazioan, Leave One Out Cross Validation teknika erabili genuen, baina azkenean "sliding-window" erabili genuenez, data kopuruarekin ez zen beharrezkoa. Horregatik, bariante modura jarri dugu Colab-aren bukaeran eta konparaketak egiteko balio dezake.

4.4.1 Leave One Out Cross Validation Arkitektura (no window)

HybridCNNRNNNoWindow klaseak sekuentzia osoak "sliding-window"-etan banatu gabe prozesatzeko diseinatutako ere duaren aldaera bat da. Geruza konboluzionalak (CNN) konbinatu ditu ezaugarri lokalak ateratzeko, bi norabideko LSTM geruzak epe luzerako mendeotasunak harrapatzeke eta sigmoide funtzio batekin fully connected irteera bat sailkapen bitarrerako; aurreko bertsioak egiten zuen moduan.

Eredu honek sarrera doitzen du sekuentzia osoak zuzenean maneiatzeko, leihoetan oinarritutako aurreprozesamenduaren beharra ezabatuz. Arkitektura hau egokia da datuak testuinguru globalean aztertzeko, leiho irristakorrak erabiltzen dituen eredu nagusiarekin alderatuz.

4.4.2 Evaluazioa (loocv)

Loocv evaluation funtzioak "Leave-One-Out" (LOOCV) balidazio gurutzatuko prozesua inplementatzen du. Prozesu horretan, esperimentu bakoitza test multzo gisa erabiltzen da, eta gainerakoek, berriz, entrenamendu multzoa osatzen dute. Lehenik, higitutako eta higitu gabeko esperimentuetan banatutako datuak kargatzen dira, eta zerrenda bakar batean konbinatzen dira iterazioa errazteko. Iterazio bakoitzean, esperimentu bat test gisa bereizten da, eta gainerakoak entrenamendu gisa. Datuak ezaugarriak antolatzen dituen funtzio baten bidez prestatzen dira, eta normalizatzen dira, datu guztiak eskala uniformearen daudela bermatzeko, eredura pasatu aurretik. Ez da leiho irristakorrik sortzen, sekuentzia osoekin lan egiten delako.

Erabilitako eredu HybridCNNRNNNoWindow bariantea da, entrenamendu begizta espezifiko bat erabiliz entrenatzen dena (train loop) tolestura bakoitzeko datuekin. Behin entrenatu ondoren, eredu dagokion test multzoan ebaluatzen da, eta iragarpenak benetako etiketekin alderatzen dira metrikak kalkulatzeko, hala nola F1 puntuazioa eta zehaztasuna. Metrika horiek tolestura bakoitzerako biltegitzen dira.

Esperimentu guztiak iteratu ondoren, lortutako metriken batez bestekoa kalkulaten da, hala nola doitasuna eta F1 score.

5 Emaizten analisia

Orain, lortutako emaitzak komentatzeko ditugu, HybridCNNRNN ereduarekin hasita.

5.0.1 HybridCNNRNN

Entrenamenduak erakusten du entrenamendu multzoaren galera nabarmen jaisten dela epoch-etan zehar, eta horrek adierazten du eredu datuetatik ikasten ari dela. Entrenamendu multzoaren F1 metrikak gora egiten du pixkanaka: lehen denboraldian 0.7247 izan zen, eta azkenekoan, berriz, 0.9829. Horrek esan nahi du doikuntza ona egin dela multzoan. Hala ere, balidazio multzoaren F1 metrika nahiko egonkor mantentzen da 0.61 inguruan, aldaketa txikiekin baina entrenamenduan hobekuntza esanguratsurik gabe. Horrek esan nahi du eredu agian gehiegi egokitzen ari dela entrenamendu datuetara, edo balidazio datuak konplexuagoak edo ez hain adierazgarriak direla.

Leiho-mailako proba guztien emaitzak normalak dira, %41,38ko zehaztasunarekin, %85,71ko recall altuarekin eta %55,81eko F1-score batekin. Horrek esan nahi du ereduak leiho gehienak positiboki sailkatzeko joera duela, eta horrek lagundu egiten diola recall altuari, baina zehaztasunak sufritu egiten du positibo faltsuen kopuru handi bat dagoelako.

Esperimentu mailan, emaitzak nabarmen hobetzen dira. Doitasuna %66.67koa da, eta recall-a %100era iristen du, horrek ereduak behar bezala identifikatzen dituela klase positiboko esperimentu guztiak adierazten du. F1-score %80 da, balio ona da, eta erakusten du eredu fidagarriagoa dela esperimentu osoen sailkapenean banakako leihoetan baino. Horren arrazoia izan daiteke esperimentu mailan majority voting banakako leihoetan egindako akatsak leuntzen laguntzen duela.

Azkenik, esperimentu mailako iragarpenek [0, 1, 1, 1] dira. Benetako etiketen [0.0, 0.0, 1.0, 1.0] aurrean erakusten du ereduak zuzen identifikatzen dituela esperimentu positiboetako bi, baina negatibo batean huts egiten duela. Horrek esan nahi du klase positiborako joera dagoela, eta datuen edo ereduaren diseinuaren desbalantze baten eragina izan liteke.

5.1 HybridCNNRNNNoWindow

Metrikak Leave-One-Out (LOOCV) balidazio gurutzatuko eredu emaitza inkonsistenteagoak dituela erakusten dute eta, oro har, baxuak direla. Metrika orokorren batez bestekoak 0.3236 F1-Score eta %43,76ko zehaztasuna erakusten du, ereduak kasu askotan datuetan patroiak behar bezala atzematea lortzen ez duela adieraziz.

Esperimentu bakoitzeko banakako emaitzek aldakortasun handia dute. Esperimentu batzuek, 10ek adibidez, metrika oso altuak lortzen dituzte (F1-Score 0.9808koa eta %96,23ko zehaztasuna), eta horrek adierazten du eredu oso ondo bete daitekeela datuak adierazgarriagoak edo sailkatzeko errazagoak direnean. Hala ere, hainbat esperimentu, hala nola 18, 1, 2, eta 3, 0.0000 F1-Score bat dute, eta horrek esan nahi du ereduak ez duela lortzen instantzia positiborik behar bezala identifikatzea kasu horietan.

Zehaztasuna handiagoa da 10 eta 9 bezalako esperimentuetan, baina kasu gehienetan baxua edo moderatua izaten da, eta horrek iradokitzen du ereduak huts egiten duela bai klase positiboa detektatzerakoan, bai benetako positiboen eta negatiboen arteko balantze egokia mantentzerakoan.

Aipatu beharra dago, argi ez dugun arrazoiaren batengatik, 15^o esperimentuak errorea ematen zuela LOOCV erabiltzean, uste dugu datuen batek NaN balioa duela baina ez dugunez arrazoiak aurkitu, entrenamendu garaian try-except egitura bat implementatu dugu eta 15^o esperimentua ez da erabiltzen. Hori ere kontuan izaten da batzbesteko metrikak kalkulatzekoan.

6 Ondorioak

Ikerlan honetan, neurona-sare konboluzionaleko eta errekurrenteko (HybridCNNRNN) eredu hibrido bat inplementatu eta ebaluatu da, baldintza espezifikoak dituzten esperimentuetako datuak sailkatzeko. Azterketan zehar, ereduaren errendimenduan alde nabarmenak ikusi ziren. Horrek zereginaren konplexutasuna eta datuak emaitzetan duten eragina nabarmentzen ditu.

Lehen fasean, eredu emaitza nahiko onak lortu zituen metrika globalei dagokienez, batez besteko 0.8-ko F1-Score batekin, gehiengoaren bozketa aplikatu ondoren. Hala ere, ereduak orokortze gaitasun mugatua erakutsi zuen, leih mailan %45 inguruko zehaztasuna lortu zuelako. Portaera hori esperimentuen arteko heterogeneotasunarengatik izan daiteke, baita klaseen arteko balizko desbalantze bati ere, ereduak hain adierazgarriak ez diren datuetan patroiak detektatzeko duen gaitasunari eragin dezakeena.

Beste alde batetik, ereduaren bariantea, ebaluazio zehatzagoa eta doituagoa egiteko inplementatuak, LOOCV balidazioan oinarritua, emaitza erratikoagoa erakutsi zuen. Ereduak errendimendu bikaina lortu zuen esperimentu batzuk gorabehera, hala nola 10 esperimentua 0.98ko F1-Score batekin, beste kasu batzuetan, emaitzak oso baxuak izan ziren, eta 0.0000 F1-Score batera ere iritsi ziren hainbat esperimentutan. Portaera horrek sarrerako datuen kalitatearen aldakortasuna eta datuen azterketa sakonagoaren garrantzia azpimarratzen ditu. Esperimentuen arteko puntuazioen desberdintasun handiak erakusten du ereduak datuen ezaugarri espezifikoekiko duen sentikortasuna.

Ondorio orokorra da, eredu hibridoak datu esperimentalen sailkapenerako irtenbide oso ona eskaintzen duen arren, bere errendimendua, neurri handi batean, datuen kalitatearen eta sendotasunaren mende dagoela. Gainera, klaseak ekilibratzeko tekniken inplementazioa eta esperimentu mota bakoitzerako ikuspegi pertsonalizatuagoa hartu behar dira kontuan, baina daukagun datu multzoa erreduzituarekin normala da.

Azkenik, ikerketa honetan etorkizunean hobekuntza gehiago egiteko bide librea usten du. Teknika asko daude zeinek ereduaren errendimendua hobetu dezakete, adibidez arreta mekanismoen inplementazioa beste batzuen artean.

7 Colab Notebook-aren esteka

Hau da Colab koadernoaren esteka