

Program

Digitalna transformacija robotiziranih tovarnih prihodnosti

DIGITOP

Rezultat D4.6:

Algoritem za zajemanje veščin operaterjev

Projekt/RRP	DIGITOP/RRP 4 - Aktivnost 4.6 Zajemanja veščin operaterjev - algoritem
Vrsta dokumenta	Poročilo o rezultatu D4.6
Verzija dokumenta	v1
Sodelujoči partnerji	IJS-E2, MET
Avtorji	Miha Glavan, Dejan Gradišar, Gregor Mlinarič, Rok Prešeren
Datum priprave	31. 07. 2025

Kazalo

1. UVOD	3
2. ZAJEM ZNANJA NA PODLAGI IDENTIFIKACIJE RELACIJ MED PROCESNIMI SPREMENLJIVKAMI.....	3
3. ZAJEM ZNANJA PREKO RAZVOJA PODPORNO-ODLOČILNIH SISTEMOV	5
3.1. RAZVOJ NAPOVEDNIH MODELOV PROIZVODNE UČINKOVITOSTI IN UPORABA OPTIMIZACIJSKIH METOD	5
3.2. PRIPOROČILNI SISTEMI NA PODLAGI GRAFOVSKIH NEVRONSKIH MREŽ	7
BIBLIOGRAFIJA	10

1. Uvod

S staranjem prebivalstva in vse bolj pogosto menjavo proizvodnega osebja postaja vedno bolj pomembno upravljanje veščin delovne sile. To področje je v osnovi široko in zajema od popisa veščin in kompetenc zaposlenih za namene določevanja ustreznih odgovornih oseb, do prenosa znanja preko ustreznih izobraževanj in dokumentacije. V tem programu zajeta tematika pa naslavlja problematiko identifikacije veščin in dobrih praks operaterjev proizvodnih procesov. Cilj je zgolj na podlagi proizvodnih podatkov sklepati na dobre pristope, omogočiti njihovo formalizacijo (zapis v obliki pravil ali modela) ter poskrbeti za njihov prenos na nove ali manj izkušene operaterje.

Identificirani in preizkušeni so bili različni pristopi. Nekateri izmed njih omogočajo zgolj enostaven vpogled med zakonitostmi med proizvodnimi parametri, medtem ko drugi, bolj napredni, omogočajo zajem in zapis relacij med proizvodnimi nastavitvami in končnim izplenom (učinkovitostjo) proizvodnega procesa. Vsak izmed pristopov ima seveda svoje prednosti in slabosti, zato je v praksi potrebno pogosto poseči bodisi po enostavnejših ali bolj kompleksnih metodah.

V tem dokumentu so predstavljeni raziskani in preizkušeni algoritmi, ki bodo uporabljeni pri izvedbi demonstracijskih pilotov. V sledečem poglavju so predstavljeni enostavnejši pristopi za identifikacijo relacij med procesnimi spremenljivkami pri čemer za analizo uporabljamo zgolj zgodovinske procesne podatke. V poglavju 3. pa so predstavljeni modelni pristopi, kjer se na podlagi zgodovinskih podatkov procesnega obratovanja razvijejo empirični modeli procesa.

2. Zajem znanja na podlagi identifikacije relacij med procesnimi spremenljivkami

Analiza vzročnosti v operativnih proizvodnih podatkih lahko pomaga ugotoviti, kateri dejavniki (npr. parametri stroja, kakovost vhodnega materiala, okoljski pogoji) povzročajo spremembe v proizvodni učinkovitosti (npr. kakovost, produktivnost, procesne napake). Obstajajo različne metode za analizo vzročnosti, odvisno od narave in količine razpoložljivih podatkov.

Pri analizi zgodovinskih podatkov za namene identifikacije ključnih zakonitosti in interakcij med procesnimi parametri in proizvodno učinkovitostjo v literaturi zasledimo številne metode. Metode za analizo kavnosti so namenske metode, ki statistično ovrednotijo vplive med proizvodnimi spremenljivkami in tudi smer interakcije med njimi. Druga skupina pristopov izhaja iz področja strojnega učenja napovednih modelov, kjer se uporabljajo za identifikacijo najbolj napovedno koristnih regresorjev (parametrov) s katerimi lahko izboljšamo natančnost napovedni empiričnih modelov [1].

Metode za izbiro regresorjev, ki se uporabljajo v literaturi lahko v splošnem razdelimo na tri večje razrede: filtrirne metode (ang. filter methods), ovojne metode (ang. wrapper methods) in vgrajene metode (ang. embedded methods). Med *filtrirne metode* štejemo tiste metode, kjer je korak izbire vhodnih spremenljivk ločen od gradnje končnega modela. Z njimi prefiltriramo potencialne vhodne spremenljivke, s čimer še pred modeliranjem izločimo za regresijo nepomembne spremenljivke. Posledično so te metode neodvisne od

uporabljene metodologije za modeliranje in večinoma temeljijo na statističnih testih in na uporabi poenostavljenih modelnih struktur. *Ovojne metode* pa neposredno izkoristijo izbrano metodologijo modela za ovrednotenje prediktivnih sposobnosti posameznega nabora vhodnih spremenljivk. Takšno izbira vhodov lahko predstavimo kot problem optimizacije strukture modela, kjer skozi optimizacijo ovrednotimo in primerjamo vse ali zgolj del vseh potencialnih kombinacij vhodnih spremenljivk. Kot optimalni niz vhodnih spremenljivk izberemo tisto kombinacijo vhodov s katero dobljeni model dosega najboljšo kvaliteto napovedi. *Vgrajene metode* za izbiro relevantnih spremenljivk so neposredno povezane z učenjem končnega modela. Že med učenjem modela so tako nepomembne spremenljivke postopoma izločene iz modela (npr. regularizacija, obrezovanje nevronske mreže, itd.). Vgrajene metode so v primerjavi z ovojnimi metodami računsko enostavnejše, saj ni potrebe po ponovnem učenju modela za vsako vhodno kombinacijo, temveč lahko že z enim učenjem modela poenostavimo njegovo strukturo in izločimo za regresijo nepomembne vhodne spremenljivke.

Pri filtrirnih in ovojnih metodah je iskanje vhodnih spremenljivk sestavljeno iz naslednjih štirih korakov: izbira potencialnih vhodov, ocena izbranih vhodov, zaustavitveni kriterij in validacija modela z izbranimi vhodi. Optimalno rešitev lahko dosežemo s preizkušanjem vseh potencialnih vhodnih kombinacij. Takšen način iskanja je računsko izjemno zahteven, zato so se uveljavile hitrejše metode, kjer postopoma dodajamo ali odstranjujemo vhode ter preverjamo spremembo kvalitete dobljenega modela. Takšne metode so lahko sekvenčne kot so vnaprejšno iskanje (ang. Forward Selection), povratna eliminacija (ang. Backward Elimination) in pa njuna kombinacija (ang. Stepwise Selection); ali pa naključni pristop iskanja vhodov, kot je na primer genetski algoritem.

Vsi zgoraj omenjeni pristopi sklepajo na zakonitosti v podatkih na podlagi takšne ali drugačne statistične analize ali identifikacije ustreznih matematičnih relacij med signali ali atributi. V primeru redkih podatkov (malo vzorcev) pa je problem identifikacije vzročnosti in ocenjevanje vplivov med spremenljivkami procesa še bolj zahtevno. V teh primerih je potrebno uporabiti prave predpostavke in orodja. Glavna težava redkih podatkov je, da omejuje moč statističnih metod pri ločevanju korelacije od vzročnosti ter vpliv šumov ter zunanjih deviacij procesa. Vendar pa lahko pomaga več tehnik, odvisno od narave redkosti (manjkajoče vrednosti, majhna velikost vzorca ali malo dogodkov). Temeljne smernice za izvajanje analize vzročnosti, kadar je na voljo zgolj malo vzročnih podatkov so:

- uporabljati je potrebno enostavnejše modele (npr. linearna regresija),
- izkoristiti je potrebno predhodno znanje,
- izkoristiti je potrebno naravne eksperimente (deviacije normalnega obratovanja, izredni dogodki).

Izbran pristop je močno odvisen od tipa problema (npr. časovne vrste / diskretni parametri) ter količine razpoložljivih podatkov. V nadaljevanju je predstavljenih nekaj pristopov, ki so bili preizkušeni tekom izvajanja programa.

Uporaba domenskega znanja – Najboljše za začetno raziskovanje in poenostavitev problema. Za ta korak je potrebno razumevanje domenskega znanja procesa na podlagi katerega se izvede prvo filtriranje možnih vplivov na sam proces.

Korelacijska analiza - omogoča hitro prepoznavanje redundantnih spremenljivk. Za vse neodvisne spremenljivke izračunamo Pearsonove korelacijske koeficiente in jih prikažemo v simetrični korelacijski matriki, kjer vsak

element predstavlja korelacijo med dvema spremenljivkama. Identificiramo močno korelirane značilke in jih odstranimo, saj ne prinašajo dodatnih informacij in lahko povzročijo težave zaradi kolinearnosti.

Izračun skupne informacije (MI - mutual information) - s to metodo lahko preverimo ali obstajajo nelinearne povezave in ocenimo informacijsko vrednost izbranih značilk. Ta analiza nam pove, v kolikšni meri posamezna značilka vpliva na ciljno spremenljivko. Ker je analiza MI računsko zahtevnejša, jo je smiselno izvesti šele po korelacijski analizi.

Parcialna korelacija - meri korelacijo med dvema spremenljivkama ob upoštevanju vpliva vseh ostalih značilk. Na ta način lahko ugotavljamo povezanost dveh značilk tudi po tem, ko odstranimo vpliv drugih značilk. Za filtriranje na osnovi te metode uporabljamo dva kriterija: absolutno korelacijo in p-vrednost. Absolutna korelacija nam pove, ali ima značilka zadostno linearno povezavo s ciljno spremenljivko, medtem ko p-vrednost določa, ali je ta povezava statistično značilna. Značilke ohranimo le, če sta izpolnjena oba pogoja.

LASSO regresija – pri gradnji napovednih modelov lahko uporabimo tudi metodo LASSO (Least Absolute Shrinkage and Selection Operator), ki poleg prilagajanja modela samodejno izvaja selekcijo značilk. Manj pomembnim značilkam priredi ničelne uteži in jih s tem izključi iz modela.

Grangerjeva vzročnost – Metoda za analizo kavzalnosti časovnih vrst podatkov. Pri tem analiza zaznava le časovno predvidljivost in pa linearne odnose med spremenljivkami. Metoda ovrednoti ali določen signal vsebuje informacije, ki omogočajo napovedovati časovno zakasnitev informacij v drugem signalu. Na podlagi zakasnitve informacij nato sklepa na kavzalnost obeh signalov.

3. Zajem znanja preko razvoja podporno-odločilnih sistemov

3.1. Razvoj napovednih modelov proizvodne učinkovitosti in uporaba optimizacijskih metod

Prvi podporno-odločilni pristop temelji na razvoju empiričnega modela zadostne natančnosti napovedi proizvodne metrike učinkovitosti ter uporabe optimizacijskih metod za preiskovanje boljših nastavitvev procesa (t.j. priporočilna akcija A_t).

Napovedni model (M_t) na podlagi trenutnih vrednosti spremenljivk stanja S_t in manipulativnih spremenljivk A_t napove spremembo mase produkta v naslednjem časovnem koraku ($Y_{increase,t+1}$). Za napoved se lahko uporabijo raznoliki modeli strojnega učenja. Za potrebe DIGITOP demonstracijskih problemov smo testirali in pripravili napovedne modele XGBoost, MLP ter linearno regresijo. Pri tem je potrebno izpostaviti, da je težko vnaprej predvideti najbolj ustrezno napovedno metodo, saj izbira temelji na tipu problema, količini razpoložljivih podatkov ter kompleksnosti relacij med vplivnimi in napovednimi parametri.

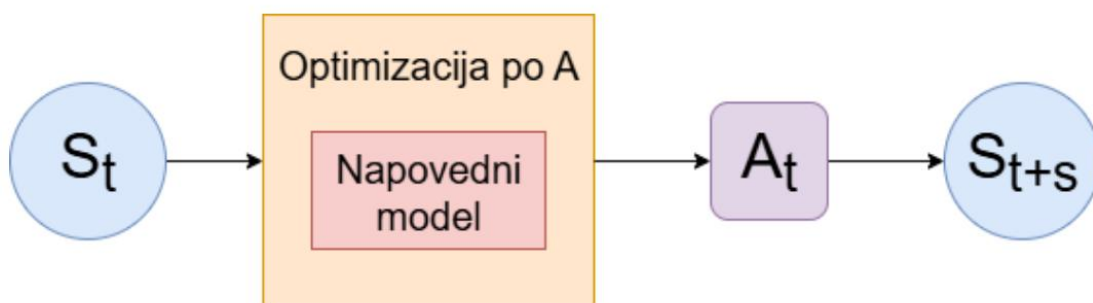
Optimizacijska metoda (Opt) poišče optimalne vrednosti akcij A_t na osnovi napovedanih izmerjenih vrednosti S_t na podlagi napovednega modela M_t . Za namene različnih problemov kriterijskih funkcij in kompleksnosti

uporabljenih modelov so bile pripravljene različne optimizacijske metode. Le-te so: gradientni spust, Adam optimizacija, Bayesovska optimizacija, CMA-ES in diferencialna evolucija.

Praktičen pristop za razvoj modelno podprtih podporno-odločitvenih sistemov je, da na podlagi urejenih zgodovinskih podatkov najprej identificiramo ustrezen napovedni model. Pri tem preizkusimo različne modelne pristope in reduciramo nabor vplivnih spremenljivk na zgolj tiste, ki so najbolj vplivne. Za slednje lahko uporabimo metode opisane v poglavju 2. Kakovost razvitega napovednega modela testiramo na neodvisnih podatkovnih setih pri čemer damo prednost robustnosti napovedi za raznolike procesne pogoje. Izogibati se je potrebno tudi prenaučenosti modelov, saj le-takrat ne zajemamo dejanskih zakonitosti v podatkih.

V naslednjem koraku definiramo optimizacijski kriterij (cenilko), s katero definiramo optimizacijski problem. V optimizacijskem kriteriju optimiramo proizvodno učinkovitost in preko mehkih omejitev vključimo tudi procesne omejitve manipulativnih spremenljivk.

Kot je to prikazano na Slika 5, lahko s takšnim pristopom v določenem odločitvenem trenutku t identificiramo priporočeno akcijo $A_{opt,t}$, na podlagi trenutnega stanja proizvodnega procesa S_t . Pri tem z optimizacijskim algoritmom invertiramo identificirano znanje formalizirano v obliki empiričnega modela in izluščimo primerno priporočilo glede na trenutno stanje proizvodnega procesa.



Slika 1: Podporno-odločilni sistem z uporabo napovednega modela proizvodne učinkovitosti in optimizacije

Slabosti in omejitve takšnega pristopa so, da so osnovne zakonitosti med proizvodnimi spremenljivkami identificirane kot črna škatla. Relacije sicer je mogoče ovrednotiti ter napovedati obnašanje sistema, vendar velika kompleksnost modela in nefizikalna naravnana struktura odločanja modela ni neposredno razložljiva. Za razložitev vplivov in identifikacijo poenostavljenih relacij se lahko uporabijo metode razložljive AI (ang. Explainable AI). Dodatni problem pa lahko predstavlja tudi že poprej omenjena prenaučenost modela ter tudi ekstrapolacija znanja na področja procesnega delovanja, kjer model ni bil naučen.

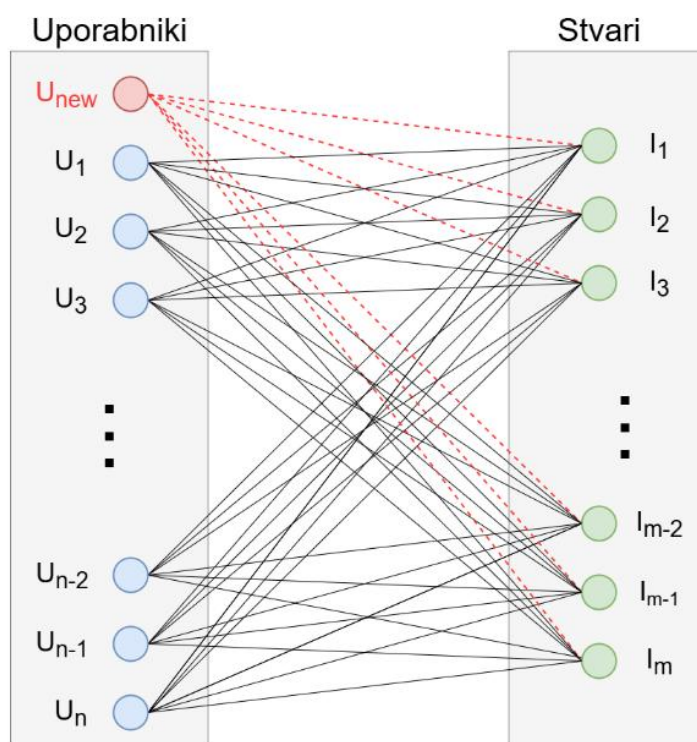
3.2. Priporočilni sistemi na podlagi grafovskih nevronske mreže

Grafovske nevronske mreže (GNN) so razred modelov globokega učenja, zasnovanih za delo z grafično strukturiranimi podatki, kjer so odnosi med entitetami prav tako pomembni kot entitete same. V kontekstu operativnih proizvodnih podatkov lahko GNN modelirajo kompleksne medsebojne odvisnosti med različnimi senzorji, stroji, proizvodnimi koraki ali procesnimi spremenljivkami.

Graf je sestavljen iz:

- vozlišč (oglišč) – entitet, kot so stroji, senzorji ali proizvodne enote
- robov – odnosov ali interakcij med temi vozlišči

Priporočilni sistem GNN je strukturiran kot bipartitni graf pri čemer vozlišča predstavljajo uporabnike in izdelke, robovi pa interakcije (npr. kliki, nakupi, ocene). Poleg tega je mogoče dodatne informacije (npr. podatke o uporabnikih) vključiti vozliščem ali robovom. Primer tipične priporočilnega sistema realiziranega z GNN modelno strukturo je prikazan na Slika 2.

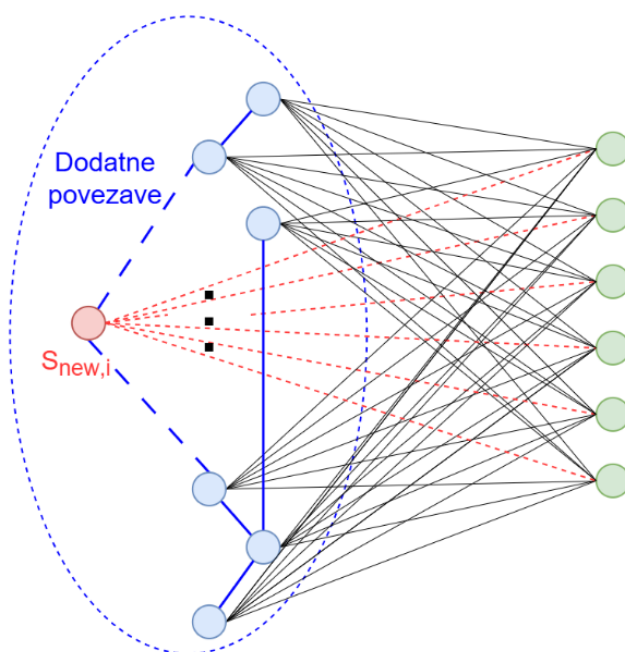


Slika 2: Struktura Grafovskega nevronskega modela za izvedbo priporočilnega sistema

Tipično predstavitev priporočilnega sistema, ki se uporablja za komercialno priporočanje izdelkov/stvari, lahko preslikamo tudi v industrijski kontekst. Pri tem leva množica vozlišč predstavlja vektorje spremenljivk stanja $S_{i,t}$ (t. j. meritve preteklih eksperimentov ob času t). Na drugi strani desna množica vozlišč predstavlja manipulativne spremenljivke, na katere lahko operater vpliva. Ta vozlišča so brez atributov. Povezave med vozlišči obeh skupin

so utežene. Vsaka povezava med vozliščem $S_{i,t}$ in manipulativnim vozliščem a_j nosi utež $w_{i,j}$, ki ustreza vrednosti manipulativne spremenljivke $a_j \in A_t$, uporabljene v eksperimentu i ob času t .

Naučene relacije med stanjem procesa, manipulativnimi spremenljivkami ter doseženo učinkovitostjo procesa so zapisane v identificirani strukturi GNN modela. Za prenos naučenega znanja na novo podobno stanje uporabimo pristop dodajanja novih vozlišč (novo stanje procesa) ter tehniko s katero napovemo robove za ta nova vozlišča (t.j. napovemo za novo stanje in različne manipulativne nastavitve kakšna bo učinkovitost procesa). Pri tem na podlagi že obstoječih povezav sklepamo na moč interakcij (učinkovitost procesa) med obema stranema bipartitnega grafa (Slika 3).

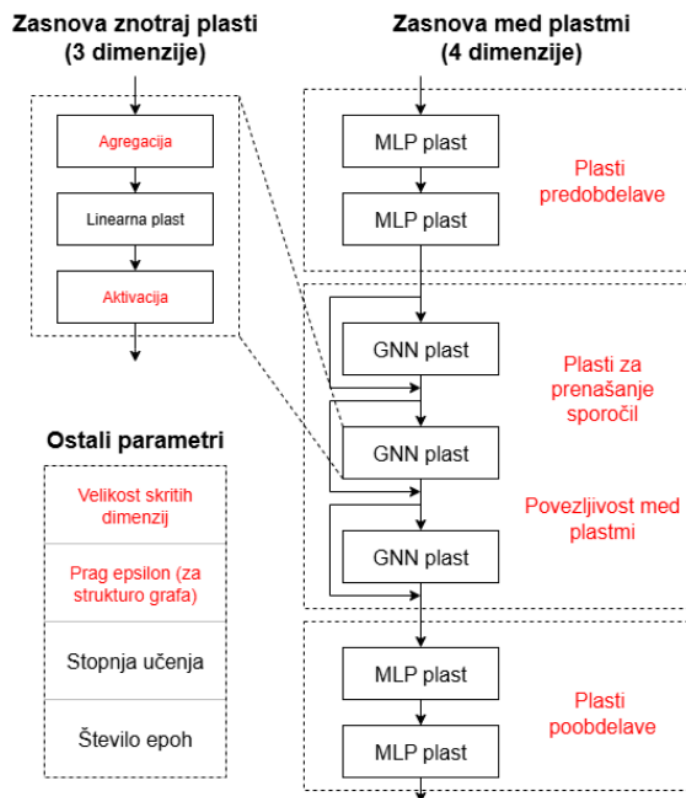


Slika 3: Dodajanje dodatnih povezav za prenos znanja in iskanje priporočil

Ključna dejavnika za učinkovitost pristopa sta izbor primerne grafovske predstavitve podatkov in ustrezne arhitekture modela. Pri izbiri ustrezne arhitekture modela je potrebno izbrati ustrezne odločitve o strukturi GNN-jev, kot je to prikazano na sliki Slika 4. Ustreden izbor teh strukturnih parametrov (t.j. hiperparametrov) pomembno vpliva na zajem znanja iz podatkov in njihov prenos na podobne primere. Pri praktičnih testih smo skladno z metodologijo, predstavljeno v [2] pripravili skripte za preiskovanje prostora možnih konfiguracij GNN modelov (ang. *design space search*). S tem smo želeli poiskati najboljše arhitekture za napovedovanje uteži povezav. Pri tem smo ključni nabor hiperparametrov pri praktičnih preizkusih omejili na naslednje parametre:

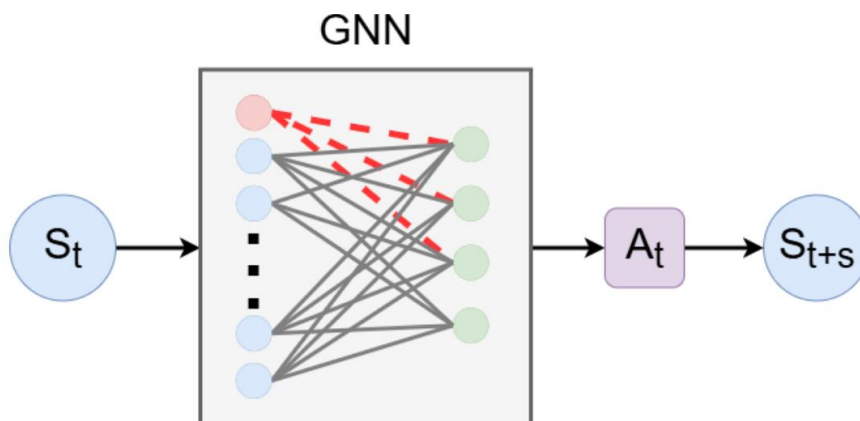
- Aktivacijska funkcija,
- Način agregacije,
- Stopnja povezanosti uporabniških vozlišč,
- Skrita dimenzija,
- Povezljivost med plastmi,
- Število plasti za prenos sporočil,

- Število plasti predobdelave,
- Število plasti poobdelave.



Slika 4: Hiperparametri grafovskih nevronske mrež

Kot je to prikazano na Slika 5, lahko z zgoraj predstavljenim pristopom v določenem odločitvenem trenutku t identificiramo priporočeno akcijo $A_{opt,t}$ na podlagi trenutnega stanja proizvodnega procesa S_t . Pri tem z vključitvijo novega vozlišča in sklepanjem na uteži tega vozlišča izluščimo primerno priporočilo glede na trenutno stanje proizvodnega procesa.



Slika 5: Priporočila na podlagi GNN modela znanja operaterjev

Slabosti in omejitve priporočilnega sistema na podlagi GNN-jev so slaba prilagodljivost, saj GNN-ji zahtevajo prenos sporočil med vozlišči, kar postane računsko drago za grafe velikega obsega (npr. veliko procesnih nastavitvev in akcij). Dodatni problem lahko predstavlja problem hladnega zagona, saj za nova področja procesnih območij ali nove nastavitve manipulativnih spremenljivk še nimamo informacij (brez ustreznih robov v grafu) zato tudi ne moremo raziskati novih še ne videnih ukrepov, temveč se omejimo na že videne rešitve ali prenos dobrih praks iz podobnih primerov. Tudi GNN-ji delujejo po principu modelov črne škatle, zato je odločanje zakaj je bilo priporočilo podano včasih težko razložiti. Tudi tukaj je mogoče uporabiti metode razložljive AI, ki omogočajo naknadno razlago in analizo odločana modelov. Velik problem predstavlja tudi pomanjkanje ustreznih podatkov, saj grafi z redkimi interakcijami lahko vodijo do slabše kakovosti priporočil saj se ne naučijo ustreznih povezav.

Bibliografija

- [1] Glavan, M., Gradišar, D., Atanasijević-Kunc, M. *et al.* Input variable selection for model-based production control and optimisation. *Int J Adv Manuf Technol* **68**, 2743–2759 (2013). <https://doi.org/10.1007/s00170-013-4840-1>
- [2] J. You, Z. Ying, and J. Leskovec, “Design space for graph neural networks,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 17009–17021, 2020.