

Razotkrivanje strukture nukleona pomoću neuronskih mreža

Marko Cvitković

Mentor: prof.dr.sc. Krešimir Kumerički

Fizički odsjek, Prirodoslovno-matematički fakultet, Bijenička 32, Zagreb

25. siječnja 2020.

1 Sažetak

U ovom seminaru promatra se ekstrakcija Comptonskih form faktora (CFF) pomoću umjetno stvorenih opservabli (udarnih presjeka) u danom kinematičkom području. CFF-ovi koji se uče na simuliranim podacima, parametriziraju se neuronskim mrežama koje su se pokazale kao jako dobra opcija zbog svojih svojstava. Računanjem RMS pogreške ekstrahiranih CFF-ova u odnosu na one iz referentnog modela, traže se najpogodnije opservable za ekstrakciju tri promatrana CFF-a. Također je pokazano da opservable *BCA* (asimetrija naboje snopa) i *TSA* (asimetrija spina mete) u kombinaciji daju najtočnije ekstrahirane CFF-ove.

2 Teorijski uvod

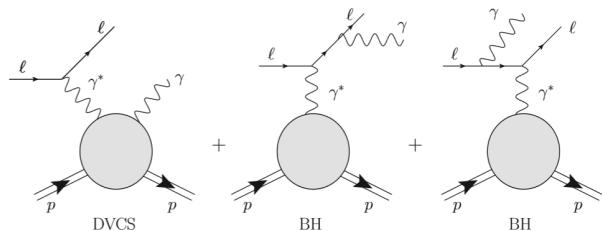
Motivacija i jedan od razloga za konstukciju novog EIC sudarivača je poznavanje generaliziranih partonskih distribucija (GPD), koje daju detaljnu sliku raspodjele kvarkova i gluona u nukleonima što je dugoročni i konačni cilj istraživanja u ovom području[1]. Postoji osam realnih GPD-ova i označavaju se velikim pisanim slovima: $H^q, E^q, \tilde{H}^q, \tilde{E}^q, H_T^q, E_T^q, \tilde{H}_T^q, \tilde{E}_T^q$. Te funkcije povezane su sa takozvanim komptonским form faktorima (CFF) preko sljedećih konvolucija[2]:

$$\mathcal{F}(\xi, t) = \sum_q e_q^2 \int_{-1}^1 dx \times \left[\frac{1}{\xi - x - i\epsilon} - \frac{1}{\xi + x - i\epsilon} \right] F^q(x, \xi, t), \quad (1)$$

$$\tilde{\mathcal{F}}(\xi, t) = \sum_q e_q^2 \int_{-1}^1 dx \times \left[\frac{1}{\xi - x - i\epsilon} - \frac{1}{\xi + x - i\epsilon} \right] \tilde{F}^q(x, \xi, t). \quad (2)$$

CFF-ovi su kompleksne funkcije i označavaju se kaligrafskim slovima i također ih je osam: $\mathcal{F} = \mathcal{H}, \mathcal{E}, \dots$ U najnižem redu doprinose $\mathcal{H}, \mathcal{E}, \tilde{\mathcal{H}}$ i $\tilde{\mathcal{E}}$. Iz danih formula vidi se da su CFF jednostavnije jer ovise o dvije varijable koje određuju kinematičko područje te funkcije: bezdimenzionalni ξ koji se može jednostavno povezati sa Bjorkenovom x varijablom x_B i Mandelstamovoj t varijabli. S druge strane GPD funkcije imaju ovisnost o još jednoj varijabli.

Srednjeročni cilj istraživanja je dobro poznavanje CFF funkcija na kojem se radi pomoću eksperimenata na JLab akceleratoru. Glavni i najpogodniji eksperiment za određivanje CFF funkcija je duboko virtualno komptonско raspršenje (DVCS) u kojem je ulazni foton virtualan i dobija se zračenjem leptona. Radi se o leptoprodukciji fotona $ep \rightarrow ep\gamma$ kojoj doprinosi i druga amplituda, tzv. Bethe-Heitler raspršenje prikazano na slici 1.



Slika 1: Leptoprodukcija realnog fotona kao koherentna superpozicija DVCS i Bethe-Heitler amplituda. Preuzeto iz [2].

U eksperimentima se mjeri udarni presjek za proces $ep \rightarrow ep\gamma$ i on je dan kao složena formula nepoznatih CFF-ova. Ključno je pitanje kako iz tih mjerenja odrediti 4 kompleksne CFF funkcije. Uvelike bi pomoglo kada bi na raspolaganju imali snopove leptona različitih naboja (elektron i pozitron) i spinova, kao i mogućnost različite polarizacije mete. Također bi pomoglo kada bi za metu mogli koristiti i proton i neutron. Na taj način dobile bi se različite opservable kojima različiti CFF-ovi različito doprinose. Neke od tih pretpostavki

nažalost još nije moguće eksperimentalno ostvariti sa željenom preciznošću. U ovom seminaru koriste se generirana mjerenja pet opservabli za ekstrakciju 3 CFF-a, ImH , ReH i $Im\bar{H}$ u svrhu saznanja koje su opservable najpogodnije za njihovu ekstrakciju. Prva opservable označava raspršenje polariziranog elektrona na nepolariziranom protonu, za što ima dosta mjerenja koja se obično onda izražavaju kao "nepolarizirani" udarni presjek σ . Zapravo se radi o zbroju spin-up i spin-down snopa i označava se sa BSS (*engl. beam spin sum*). Zanimaju nas nulti i prvi kosinusni harmonici koje promatramo kao odvojene opservable, BSSw0 i BSSw1. Analogno, "polarizirani" udarni presjek predstavlja njihovu razliku i označava se sa BSD (*engl. beam spin difference*). Koriste se još TSA (*engl. target spin asymmetry*) i BCA (*engl. beam charge asymmetry*).

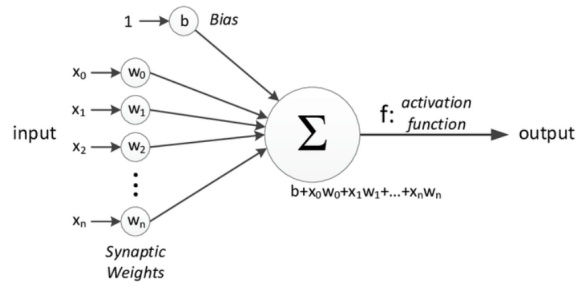
Postoji više metoda izvlačenja (ekstrakcije) nepoznatih CFF-ova iz navedenih opservabli. Neke od njih su jako primitivne i nepouzdanе kao što je tzv. lokalna ekstrakcija gdje se odrede numeričke vrijednosti nekih od CFF-ova za fiksne parametre ξ i t . U drugoj varijanti lokalne ekstrakcije, teorijski se odrede minimumi i maksimumi vrijednosti svih CFF-ova što dovodi do konvergencije i relativno preciznih vrijednosti za ImH koji je dominantni doprinos za većinu opservabli, dok su ostale vrijednosti relativno nepouzdanе.

Nama je interesantno dobiti globalnu ekstrakciju nasuprot lokalne. Jedna od opcija je pomoću teorijskih saznanja pretpostaviti funkcijski oblik za svaki od CFF-ova sa po nekoliko slobodnih parametara koje naučimo prilagodbom na mjerenja opservabli, za sve ξ i t istovremeno. To funkcionira dobro, no postavlja se pitanje pristranosti koja se unosi u model samim odabirom funkcijskih oblika. Tim odabirom značajno smanjujemo funkcijski prostor jer funkcija ovisi o dvije kinematičke varijable na relativno nepoznat način, a za konačni cilj (poznavanje GPD-ova) taj bi problem postao još i veći. Upravo je to motivacija zbog koje za ovu svrhu pribjevamo neuronskim mrežama koje parametriziraju proizvoljne CFF funkcije bez da unose pristranost.

3 Metoda: neuronske mreže

Pojam „neuronska mreža” potječe od pokušaja pronalaska matematičke reprezentacije obrade informacija u biološkim sustavima (McCulloch i Pitts 1943.) pa se tako i danas koristi za široki spektar algoritama za prepoznavanje uzoraka modeliranih po uzoru na ljudski mozak. Naime, biološki neuroni su fundamentalne jedinice mozga i živčanog sustava odgovorne za primanje senzorskih informacija iz okoline preko dendrita, a potom i za obradu istih.

Prvi korak bila je pojava algoritma perceptrona koji se koristi za binarnu klasifikaciju, a prikazan je na slici 2.



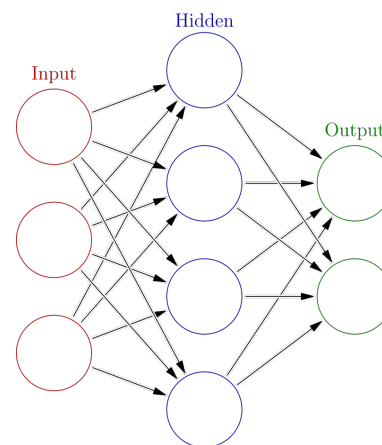
Slika 2: Perceptron. Preuzeto iz [9].

Sa gornje slike vide se glavni njegovi elementi. Ulazi x_0, x_1, \dots i x_n predstavljaju značajke jednog primjera, dok je b takozvani „bias” odnosno pristranost koji služi da decizijska granica u prostoru značajki ne prolazi nužno kroz ishodište. Svaka od tih značajki ulazi u model sa jednom od $n + 1$ težina $w_0 = 1, w_1, \dots, w_n$ koje su proporcionalne važnosti pripadne značajke:

$$a = b + \sum_{j=1}^n w_j \cdot x_j \quad (3)$$

Djelovanjem aktivacijske funkcije na skalarni produkt značajki sa težinama dobijamo predikciju za dani primjer. Aktivacijska funkcija kod perceptrona je step funkcija, zbog čega je perceptron linearan model. Optimizacija se vrši tako da se predikcija primjera uspoređuje sa danim oznakama skupa za učenje i težine se ažuriraju ukoliko oznaka ne odgovara predikciji sve dok se ne nađe vektor težina za koji su sve predikcije jednake stvarnim oznakama primjera.

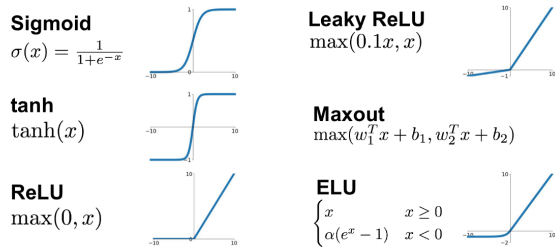
Neuronske mreže su u biti višeslojni perceptroni, ulazni čvorovi kombiniraju se sa pripadnim težinama na opisani način u više čvorova u takozvanom skrivenom sloju koji sada služi kao ulaz za sljedeći sloj itd. Primjer arhitekture umjetne neuronske mreže dan je na slici 3.



Slika 3: Umjetna neuronska mreža. Preuzeto iz [8].

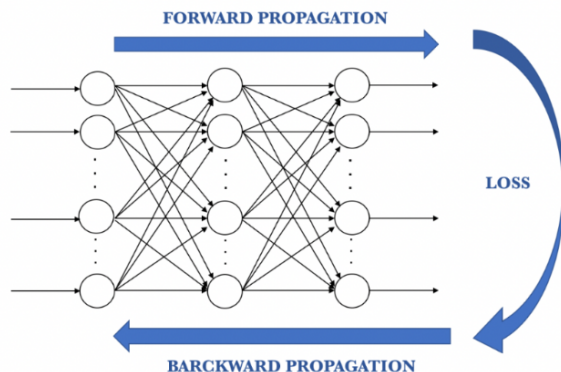
Ipak bitna razlika između perceptrona i neuronske

mreže je u aktivacijskoj funkciji. Kod perceptrona, kao što je spomenuto, radi se o step funkciji koja odlučuje je li čvor „aktiviran” ili ne. Kada bi imali više slojeva, izlaz bi i dalje bio linearna kombinacija izlaza. Takav model ima jako malenu složenost i nema kapacitet naučiti i modelirati većinu složenih podataka. U neuronskim mrežama koriste se nelinearne aktivacijske funkcije koje se primjenjuju na skrivene slojeve, zbog čega se neuronske mreže smatraju univerzalnim aproksimatorima.



Slika 4: Nelinearne aktivacijske funkcije. Preuzeto iz [10].

Na slici poviše, dane su često korištene nelinearne aktivacijske funkcije. Broj skrivenih slojeva i broj čvorova u slojevima ovisi o zadanom problemu i području u kojem se algoritam primjenjuje (duboko učenje, obrada prirodnog jezika, računalni vid ...). Neuronska mreža parametrizirana je sa težinama i „biasima” svakog neurona koje je potrebno što bolje naučiti kako bi neuronska mreža obavljala traženu zadaću. Učenje se odvija uz pomoć algoritma propagacije unatrag (*engl. backpropagation*). Kod tipa učenja koje nazivamo nadzirano, u skupu primjera za učenje imamo značajke i pripadne oznake kod klasifikacije ili vrijednosti ciljane varijable kod regresije. U prolasku mrežom unaprijed, dobijemo predikciju iz koje računamo pogrešku prema funkciji pogreške prikladnoj za problem koji rješavamo.



Slika 5: Ciklus učenja neuronske mreže. Preuzeto iz [9].

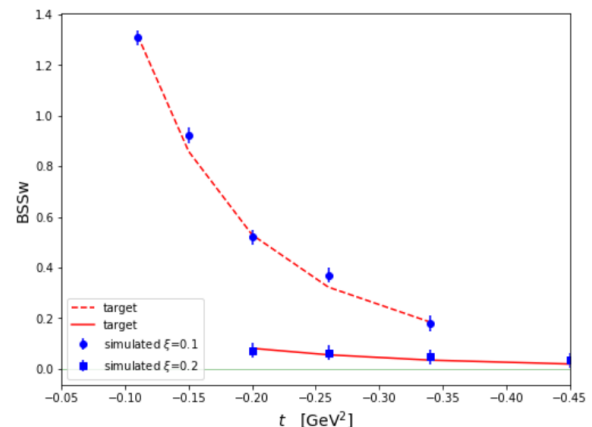
Računaju se gradijenti funkcije pogreške po svim parametrima, odnosno težinama koje se onda ažuriraju tako da se od njih oduzima pripadni gradijent pomnožen

sa hiperparametrom (treba ga također optimizirati unakrsnom provjerom) kojeg nazivamo stopa učenja (*engl. learning rate*). Pokazuje se da se pomoću ulančanog deriviranja dijelovi gradijenata dalje od ulaza mogu iskoristiti u slojevima bliže ulazu zbog čega je algoritam puno efikasniji i objašnjava samo ime.

4 Rezultati i diskusija

Pomoću koda u programskom jeziku Python koji je priredio mentor ponavlja se sljedeća procedura:

- Zamislimo da znamo sve CFF funkcije koje nas zanimaju, odnosno koristimo referentni poznati Goloskokov-Kroll[4] u svojoj pojednostavljenijoj formi gdje pretpostavljamo da samo CFF-ovi koji nas zanimaju imaju nezanemarive doprinose za razliku od standardnog GK modela koji u obzir uzima svih 8 CFF-ova koji čine tzv. „*leading-twist order*”.
- Generiramo vrijednosti udarnih presjeka (opservabli) za CFF funkcije u nekom odabranom kinematičkom rasponu sa simuliranim neodređenostima što zovemo simuliranim podacima (*engl. mock data*). Korisno je napomenuti kako je korišteni kinematički raspon u ovom radu definiran sa vrijednostima ξ koji je između 0.1 i 0.25 i t koji je između -0.45GeV^2 i -0.1GeV^2 . Za spomenuto generiranje simuliranih podataka se koriste poznate BMK formule u kojima je sadržana složena ovisnost opservabli o Comptonkim faktorima[6]. Na slici 6. prikazane su neke vrijednosti simuliranih podataka iz mjerenja za varijablu BSSw1 i njihova usporedba sa podacima koji odgovaraju referentnom modelu.



Slika 6: Prikaz simuliranih podataka za opservablu BSSw1 u ovisnosti o kinematičkim varijablama

- Ekstrahiramo željene CFF-ove koje smo parametrizirali neuronskim mrežama sa kinematičkim va-

rijablama na ulazu, tako da uspoređujemo udarne presjeke izračunate pomoću dobijenih vrijednosti CFF-ova sa imitacijama mjerenih vrijednosti (simuliranim podacima). Prvi korak je takav da su težine postavljene na neke nasumične vrijednosti, te se iste ažuriraju tako da se dobijena pogreška propagira unatrag na opisani način što dovodi do smanjenja pogreške.

- Dobijene CFF-ove uspoređujemo sa poznatim CFF-ovima (dobijenim pomoću referentnog modela) te promatramo je li čitava procedura konzistentna.

Valja napomenuti kako svaku ekstrakciju provodimo tako da se uzimaju neuronske mreže za koje je vjerojatnost uz pripadajući χ^2 veća od praga 0.05. Postoji parametar *nnets* koji određuje koliko takvih mreža nam je potrebno. Tako se propagiraju eksperimentalne nepouzdanosti i kao rezultat prilagodbe neuronskih mreža dobiju se vrpce koje odgovaraju nepouzdanosti (standardne devijacije oko prosječne vrijednosti od CFF-ova ekstrahiranih *nnets* puta) kao što je prikazano na slici 7. gdje je iz svih pet opservabli istovremeno uspješno ekstrahirano tri promatrana CFF-a.

Prva dva stupca prikazuju presjek plohe Compton-skog form faktora po ξ za $t = -0.2GeV^2$ gdje je u drugom stupcu sa zelenom označeno kinematičko područje koje promatramo dok je izvan dana ekstrapolacija. U trećem i četvrtom stupcu prikazan je presjek po osi $-t$ za dvije vrijednosti parametra ξ koje su 0.12 i 0.22 respektivno.

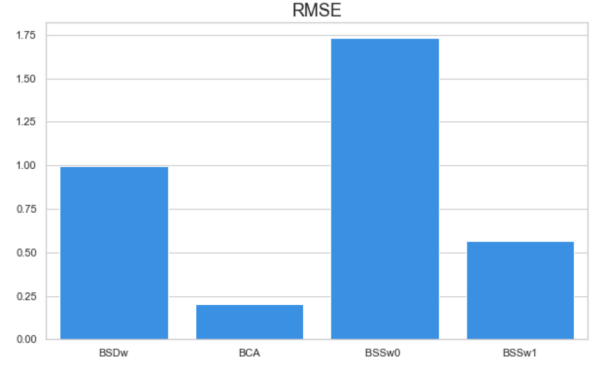
Cilj je u tako stvorenom okruženju pokazati jesu li neke opservable ili kombinacija opservabli više ili manje pogodne za ekstrakciju CFF-ova. To smo se odlučili na način da za razne kombinacije mjerenja tj. simuliranih opservabli pokušamo uspješno ekstrahirati po jednu ili više CFF-ova te onda računati pogrešku u odnosu na referentni model.

Koristi se korijen iz kvadratne srednje pogreške (*engl. root mean square error*) koja je definirana kao:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2}, \quad (4)$$

gdje je N broj primjera odnosno dimenzija vektora Y opaženih vrijednosti i vektora \hat{Y} predviđenih vrijednosti.

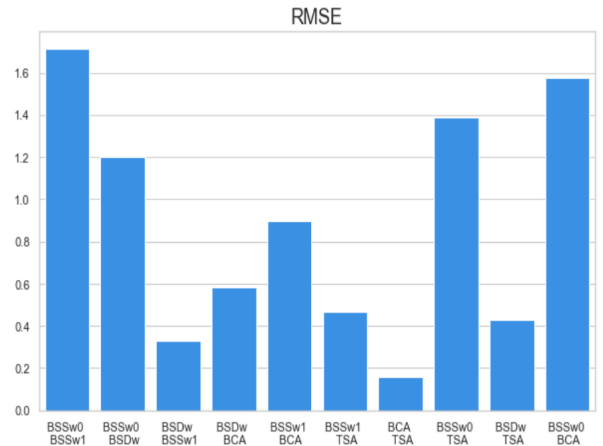
Početni korak bio je ekstrahirati iz svih 5 opservabli po 1 CFF, dominantni doprinos. RMSE vezan uz opservablu je RMSE dominantnog doprinosa ekstrahiranog pomoću te varijable. Situacija sa ovim kombinacijama prikazana je na slici 8.



Slika 8: RMSE dominantnog doprinosa CFF-a pri ekstrakciji iz jedne opservable

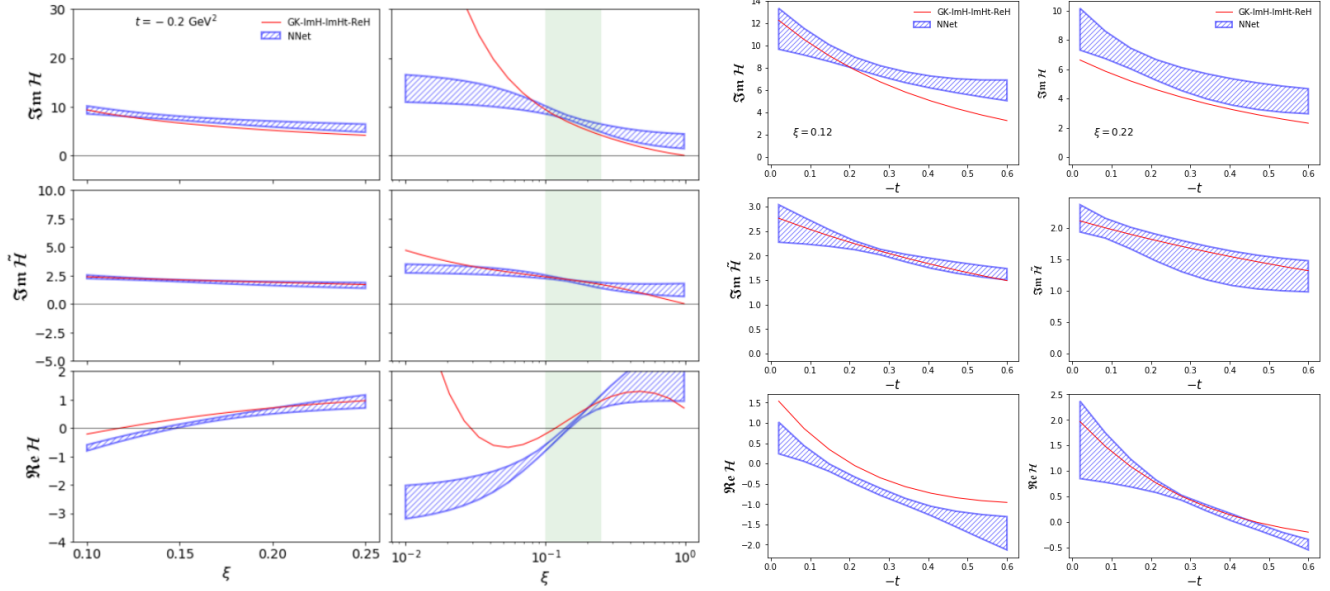
Kao što se vidi, na slici nedostaje situacija sa opservablom TSA. Po teorijskim saznanjima iz TSA bi bilo lakše izvući *ImH* i *ImHt* skupa nego ijedan CFF posebno jer iako je *ImHt* nominalno dominantan, *ImH* se ne može zanemariti. Međutim ni to mi nije pošlo za rukom uz mnoge pokušaje raznim brojem skrivenih slojeva što daje složenije modele, kao i promjenom broja čvorova u skrivenom sloju. Inače se dobijaju rezultati konzistentni s teorijom. Primjerice, za BSDw dominantan je *ImH* koji je nezanemariv u velikoj većini slučajeva, dok je za BCA dominantan *ReH*.

Nadalje, promatrala su se i RMSE pri ekstrakciji po 2 dominantna CFF-a uz „mjerenja” svih kombinacija po dvije opservable. U ovom slučaju ukupni RMSE koji odgovara ekstrakciji uz kombinaciju dvije opservable izračunat je kao prosjek dvije pogreške uspješno ekstrahiranih dvaju CFF-ova. Rezultati su prikazani na slici 9.



Slika 9: Prosječni RMSE dvaju CFF-ova pri ekstrakciji iz dvije opservable

Vidi se da kombinacija opservabli *BCA* i *TSA* daje najbolje rezultate. Slijede *BSDw* i *BSSw1*, zatim *BSSw1* i *TSA* te *BSDw* i *TSA*.

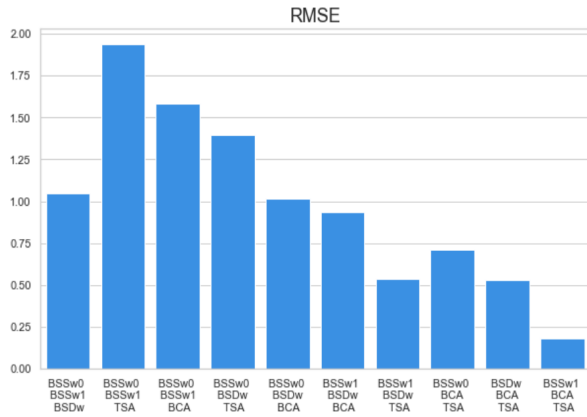


Slika 7: Prikaz tri istovremeno ekstrahirana CFF-a

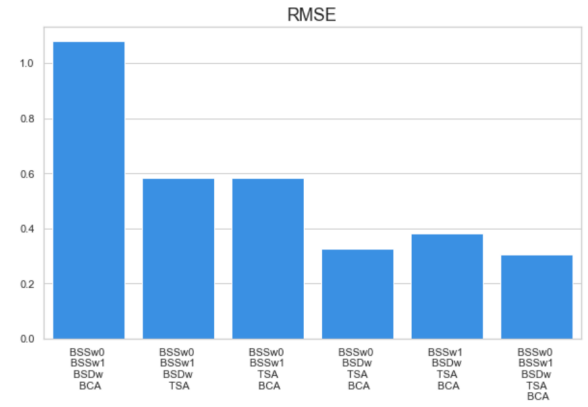
Zatim su se promatrale sve moguće kombinacije od po 3 opservable te se radila analogna procedura. Ovoga puta za dana mjerenja mogli smo ekstrahirati sva tri CFF-a za svaku kombinaciju opservabli te je shodno tome prikazan prosjek RMS pogrešaka za po tri ekstrahirana CFF-a. Rezultati su prikazani na slici 10.

Ponovno primjećujemo da se među kombinacijama s najmanjom prosječnom pogreškom najčešće pojavljuju kombinacija *BCA* i *TSA* uz moguće treće varijable.

Ostaju još slučajevi sa svim mogućim kombinacijama od 4 opservable kojih ima 5 kao i jedan slučaj gdje ekstrahiramo sve CFF-ove pomoću svih opservabli koje su nam na raspolaganju. Tih 6 kombinacija prikazano je zajedno na slici 11. gdje vrijedi ista analogija ko prije: prikazana je prosječna RMS pogreška pri ekstrakciji sva 3 promatrana CFF-a.



Slika 10: Prosječni RMSE tri CFF-a pri ekstrakciji iz tri opservable



Slika 11: Prosječni RMSE tri CFF-ova pri ekstrakciji iz četiri i pet opservabli

I po treći put u ovom istraživanju se vidi da se kombinacija opservabli *BCA* i *TSA* pojavljuje u sve tri kombinacije sa najmanjom prosječnom pogreškom što ukazuje da su baš te dvije opservable u kombinaciji najpogodnije za ekstrakciju tri promatrana Comptonska form faktora. To u principu reproducira neke od rezultata koje je dobio kolega Čorić u svom diplomskom radu [11] koristeći za to metode dubokog učenja, odnosno neuronske mreže u modernom paketu *Tensorflow* sa puno većim brojem skrivenih slojeva i neurona u njima nasuprot arhitekture jednog skrivenog sloja sa najčešće 7 neurona kakva se koristila u ovom radu.

5 Zaključak

U ovom istraživačkom seminaru prikazan je pregled kroz najkvalitetniju metodu ekstrahiranja Comptonskih form

faktora, a to je pomoću neuronskih mreža iz umjetno generiranih podataka. Također je pokazano da opservable BCA i TSA u kombinaciji daju najtočnije ekstrahirane CFF-ove.

Literatura

- [1] Accardi, A., Albacete, J.L., Anselmino, M. et al, *Electron Ion Collider: The Next QCD Frontier*, Eur. Phys. J. A (2016) 52: 268.
- [2] Krešimir Kumerički, Simonetta Liuti, and Herve Moutarde, *GPD phenomenology and DVCS tting. Entering the high-precision era*. European Physical Journal A, 52:157, June 2016.
- [3] Kresimir Kumerički, Dieter Mueller, Andreas Schafer, *Neural network generated parametrizations of deeply virtual Compton form factors*, A. J. High Energ. Phys. (2011) 2011: 73. [https://doi.org/10.1007/JHEP07\(2011\)073](https://doi.org/10.1007/JHEP07(2011)073)
- [4] S. V. Goloskokov and P. Kroll, *The role of the quark and gluon GPDs in hard vector-meson electro-production*, Eur. Phys. J. C53 , 2008, 367–384
- [5] Burkert, V.D., Elouadrhiri, L. Girod, F.X., *The pressure distribution inside the proton.*, Nature 557, 396–399 (2018) doi:10.1038/s41586-018-0060-z
- [6] A.V. Belitsky, D. Müller, A. Kirchner, *Theory of deeply virtual Compton scattering on the nucleon*, Nuclear Physics B, 629, May 2002, 323-392
- [7] Kresimir Kumerički, Dieter Mueller, Andreas Schafer, *Parametrizing Compton form factors with neural networks*, Nuclear Physics B, 222-224, January 2012, 199-203
- [8] <https://en.wikipedia.org/wiki/Artificial-neural-network>
- [9] <https://towardsdatascience.com/introduction-to-artificial-neural-networks-ann-1aea15775ef9>
- [10] <https://towardsdatascience.com/complete-guide-of-activation-functions-34076e95d044>
- [11] I. Čorić, *Istraživanje kvarkovsko-gluonske strukture protona pomoću strojnog učenja*, 2019 (Diplomski rad). Dostupno na: <https://urn.nsk.hr/urn:nbn:hr:217:856388>