

INFORME PROGRÉS 1

1 INTRODUCCIÓ:

En aquest informe de progressió, s'abordaran els objectius que s'han desenvolupat durant les dues primeres fases del projecte, la fase de formació i la fase d'avaluació. Es durà a terme una avaluació de l'eficàcia en l'assoliment dels objectius, així com del contingut treballat fins al moment. Finalment, es proporcionarà un apartat de conclusions que inclourà suggeriments per a possibles millores en la pròxima etapa del projecte. Cal destacar que en aquest informe, l'objectiu principal serà descriure i detallar les tasques i activitats desenvolupades durant les dues fases del projecte, sense abordar de forma exhaustiva la presentació dels resultats concrets.

2 ASSOLIMENTS:

En aquest apartat es presenten els objectius treballats durant les dues primeres fases del projecte, així com els mitjans i les tasques emprades per a la seva consecució. S'ha optat per representar aquesta informació de forma clara i concisa mitjançant una taula resum, que permeti visualitzar de manera ràpida i efectiva els assoliments del projecte fins al moment.

FASE DE FORMACIÓ		
Objectiu	Metodologies empleades	Resultats
O1	<p>Per a l'assoliment del primer objectiu, que consistia en desenvolupar un coneixement profund dels fonaments teòrics i pràctics de les XN i la simulació de DM, s'han emprat diverses metodologies.</p> <p>En el cas de la simulació de DM, s'han realitzat diverses tutories amb el meu tutor extern Jordi, on hem discutit a fons el tema i he pogut resoldre dubtes i aprofundir en conceptes clau. A més a més, s'ha realitzat recerca pròpia, llegint treballs que em van ser enviats pel tutor i també articles, vídeos i altres fonts de documentació que aborden aquest tema en profunditat.</p> <p>Per a l'assoliment del coneixement en XN s'han emprat diverses vies. D'una banda, les assignatures relacionades amb el tema, com ara APC, han estat fonamentals per a la comprensió dels fonaments teòrics. D'altra banda, s'han realitzat reunions i trucades amb estudiants de carreres més especialitzades en el tema, com ara IA o MatCAD, que han aportat una perspectiva diferent i enriquidora. Finalment, s'ha emprat la recerca i la lectura de treballs que tracten els temes que s'han anat trobant durant la lectura dels treballs de DM.</p>	<p>S'ha aconseguit desenvolupar un coneixement profund dels fonaments teòrics i pràctics de les Xarxes Neuronals i la simulació de DM. Això ha permès iniciar el treball del projecte amb més facilitat, així com entendre millor els objectius a aconseguir en les fases posteriors. Encara que no es disposi de resultats concrets a presentar per aquest objectiu, s'ha realitzat un informe inicial que mostra l'estat actual del projecte i les accions previstes per a la seva consecució.</p>

O2	<p>Per al desenvolupament d'una comprensió crítica dels avantatges i les limitacions de les XN en la simulació de DM, comparant-les amb altres tècniques i abordaments existents, s'han realitzat tutories amb el tutor extern Jordi per discutir aquest tema. A més, s'han fet lectures d'articles científics i treballs relacionats que presenten comparacions entre les metodologies tradicionals i les noves que apliquen les Xarxes Neuronals.</p>	<p>S'ha detectat que les XN presenten avantatges significatius en l'aplicació de tècniques de DM que involucren grans quantitats de dades, ja que són capaces de processar aquestes dades de forma eficient i eficaç. També s'ha identificat que les XN són una eina útil per a l'extracció de característiques, ja que poden identificar patrons complexos que les tècniques tradicionals no poden detectar.</p> <p>No obstant això, també s'ha detectat que les XN presenten algunes limitacions en la seva aplicació en DM, com ara la necessitat d'una gran quantitat de dades per al seu entrenament i la dificultat per interpretar els resultats obtinguts. Així mateix, s'ha comprovat que no sempre són la millor opció en casos en què les dades no són prou complexes o no es disposa d'una gran quantitat de mostres.</p>
----	---	---

FASE D'AVALUACIÓ		
Objectiu	Metodologies empleades	Resultats
O3.1	<p>Per a l'objectiu de l'estudi de les bases de dades utilitzades en els treballs més recents del camp de la DM que apliquen XN, la metodologia emprada va consistir principalment en la lectura detallada de la documentació i dels papers publicats dels treballs TorchMD^[1] i SchNetPack^[2]. Això va incloure una anàlisi exhaustiva dels conjunts de dades utilitzats en aquests treballs i una comparació de les seves característiques i limitacions. També es van estudiar altres fonts de informació relacionades amb les bases de dades de DM i XN per ampliar el coneixement en aquest àmbit.</p>	<p>Després de llegir la documentació i els papers publicats dels treballs TorchMD i SchNetPack, s'ha observat que aquests treballs fan bases de dades com ara la QM9 per a TorchMD i de nou la QM9^[3] i la MD17^[4] per a SchNetPack entre d'altres.</p> <p>Per tant, es pot concloure que la QM9 és una base de dades amplament utilitzada en els treballs més recents de DM que apliquen XN, i serà de gran utilitat donat que facilitarà el treball a la hora de comparar els models</p> <p>(Entro en més detall sobre les bases de dades utilitzades en l'informe de el "State of the art")</p>
O3.2	<p>(Comparteix metodologia amb els objectius O3.1 – O3.2 i O3.3.)</p>	<p>En els papers publicats de s'afirma que SchNetPack proporciona dos possibles models de representació (w)ACSF (description-based) i SchNet (end-to-end). En el cas de TorchMD pren exactament aquest mateix sistema per representar els sistemes.</p>

		<p>Donada la poca varietat de sistemes de representació s'ha realitzat recerca extra en altres treballs. En aquest cas DeePMD-kit^[5] s'utilitza una combinació entre un mètode description-based i un end-to-end.</p> <p>(Tots els mètodes mencionats estan explicats detalladament al document de "State of the art")</p>
O3.3	(Comparteix metodologia amb els objectius O3.1, O3.2 i O3.3.)	<p>Ambdós treballs comparteixen la mateixa arquitectura la qual es flexible i adaptable a cada cas. On difereixen els dos treballs es alhora de fer la inferència i l'entrenament on TorchMD fa us de les eines disponibles al framework de PyTorch lightning^[6].</p> <p>(Entro en mes detall sobre l'estructura en l'informe de el "State of the art")</p>
O3.4	<p>Per a aquest objectiu, s'ha buscat instal·lar els dos programes de xarxes neuronals, TorchMD i SchNetPack, per a poder fer proves i comparacions entre els models. S'ha intentat seguir les instruccions proporcionades a la documentació oficial de cada programa per instal·lar-lo en un entorn de treball adequat.</p> <p>S'ha començat amb la instal·lació de SchNetPack, i malgrat els problemes que han sorgit durant aquest procés, s'han realitzat algunes proves amb les demos que s'inclouen en el programa. No obstant això, no s'ha pogut procedir amb la instal·lació de TorchMD i, per tant, no s'han pogut dur a terme les comparacions amb SchNetPack.</p> <p>Així doncs, la metodologia s'ha centrat principalment en la instal·lació i proves del SchNetPack, amb l'esperança de poder realitzar comparacions amb TorchMD en un futur.</p>	<p>Els resultats obtinguts en aquest objectiu han permès iniciar la prova de funcionament del codi SchNetPack. S'ha pogut importar les dades de la base de dades QM9 i transformar-les en el format necessari per a la representació mitjançant el model SchNet. Així mateix, s'ha definit l'estructura del model SchNet, incloent el mètode de representació i les capes de predicció.</p> <p>A més, s'ha realitzat un entrenament utilitzant PyTorch Lightning i s'ha dut a terme una prova de predicció per avaluar el model SchNet creat. Malgrat els problemes amb la instal·lació i l'execució del codi, els resultats obtinguts han permès una comprensió més profunda dels processos involucrats en el funcionament intern del framework.</p> <p>(El jupyter notebook de la demo es troba en el GitHub del projecte en la carpeta de codi.)</p>

3 CONCLUSIONS:

En primer lloc, s'ha aconseguit un bon coneixement teòric dels fonaments de les XN i la seva aplicació en el camp de la simulació de DM. S'ha explorat l'estat de l'art en aquest àmbit, destacant les avantatges i limitacions de les XN en comparació amb altres tècniques i abordatges existents.

Així mateix, s'ha realitzat una investigació sobre les bases de dades més comunes en els treballs més recents de DM que apliquen XN. S'ha identificat la QM9 com una base de dades comuna que s'utilitza en molts d'aquests treballs i que pot ser útil per a la comparació de models.

Pel que fa als resultats de les proves amb el codi, s'ha aconseguit instal·lar i executar amb èxit el SchNetPack i provar una demo que inclou l'import de dades de la base de dades QM9, la transformació de les dades al format necessari per iniciar la representació mitjançant el model SchNet, la definició de l'estructura del model i la realització d'un entrenament amb PyTorch Lightning. S'ha comprovat que el model SchNet pot ser útil per a la predicció de propietats moleculars.

És important reconèixer la desviació del planning i establir un pla per compensar-lo. És comprensible que hi hagi circumstàncies imprevistes que puguin afectar el pla inicial. En aquest cas, es pot planificar una reorganització del temps i les tasques per assegurar que els objectius inicials es compleixin en el temps establert.

Per compensar la desviació del planning, es considerarà l'opció de dedicar més temps a les tasques que s'han quedat endarrerides i si es necessari reduir el temps de les tasques menys prioritàries per assegurar que es compleixin els objectius establerts en el pla original.

4 OBSERVACIONS:

Durant el testing del model SchNetPack, s'ha descobert que una de les possibles millores que s'havia plantejat per millorar el model ja estava implementada en la versió més recent del mateix. Es tracta de la implementació de Python Lightning al model, que ja es troba disponible a la versió 2 del SchNetPack^[7]. Aquesta descoberta és important perquè ens indica que el desenvolupament del model està evolucionant i s'estan realitzant millores per optimitzar-lo i augmentar-ne la seva eficiència. Això ens motiva a estar atents a futures versions i a seguir avaluant el model SchNetPack en la recerca de noves oportunitats per millorar els resultats obtinguts en el camp de la dinàmica molecular.

5 BIBLIOGRAFIA:

- [1] - K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller. **SchNetPack: A Deep Learning Toolbox For Atomistic Systems**. *Journal of chemical theory and computation*, 2019, 15, 448-455.
- [2] - Stefan Doerr, Maciej Majewski, Adrià Pérez, Andreas Krämer, Cecilia Clementi, Frank Noe, Toni Giorgino, and Gianni De Fabritiis. **TorchMD: A Deep Learning Framework for Molecular Simulations**. *Journal of chemical theory and computation*, 2021, 17, 2355–2363.
- [3] - MD17 (Molecular Dynamics 17): <https://paperswithcode.com/dataset/md17> (accessed 5/3/2023).
- [4] - Molecular Property Prediction on QM9: <https://paperswithcode.com/sota/molecular-property-prediction-on-qm9> (accessed 05/3/2023).
- [5] - Takeru Miyagawa, Kazuki Mori, Nobuhiko Kato, Akio Yonezu. **Development of neural network potential for MD simulation and its application to TiN**. *Computational Material Science*, 15 April 2022, 111303.
- [6] - Falcon, W. PyTorch Lightning. 2019, GitHub. <https://github.com/PyTorchLightning/pytorch-lightning> (accessed 24/3/2023).
- [7] - Kristof T. Schütt, Stefaan S. P. Hessmann, Niklas W. A. Gebauer, Jonas Lederer, Michael Gastegger; **SchNetPack 2.0: A neural network toolbox for atomistic machine learning**. *J. Chem. Phys.* 14 April 2023; 158 (14): 144801