

Պաշտոնական ընդդիմախոսի կարծիք

Վահագն Նորիկի Ալթունյանի «Մեքենայական ուսուցման և բաշխված հաշվարկային մոտեցումներ քվանտային քիմիական տվյալների ստեղծման և մոլեկուլային հատկությունների կանխատեսման համար» թեմայով Ե.13.05 «Մաթեմատիկական մոդելավորում, թվային մեթոդներ և ծրագրերի համալիրներ» մասնագիտությամբ տեխնիկական գիտությունների թեկնածուի գիտական աստիճանի հայցման ատենախոսության վերաբերյալ:

Ժամանակակից դեղագործության և նյութագիտության ոլորտներում նոր մոլեկուլների հայտնաբերումը և դրանց հատկությունների կանխատեսումը մնում է առանցքային խնդիր: Քիմիական տարածության հսկայական ծավալը դարձնում է փորձարարական **հետազոտությունները գործնականում անհնար: Միննույն ժամանակ, ab initio** մեթոդները, մասնավորապես խտության ֆունկցիոնալի տեսությունը (DFT), թեև տրամադրում են բարձր ճշգրտություն, պահանջում են զգալի հաշվողական ռեսուրսներ:

Հատկապես կարևոր է նշել, որ բարձրորակ քվանտային քիմիական տվյալները հանդիսանում են անփոխարինելի հիմք նեյրոցանցային պոտենցիալների (Neural Network Potentials, NNP) մշակման համար: Այս պոտենցիալները հնարավորություն են տալիս իրականացնել մոլեկուլային դինամիկայի սիմուլացիաներ DFT-ի ճշգրտությանը մոտ մակարդակով, բայց հազարավոր անգամ ավելի արագ: Սա բացում է նոր հնարավորություններ երկարատև կենսաբանական պրոցեսների մոդելավորման, դեղանյութերի նախագծման և նոր նյութերի հայտնաբերման համար:

Այս համատեքստում, Վ.Ն. Ալթունյանի ատենախոսական աշխատանքը, որը նվիրված է մեքենայական ուսուցման և բաշխված հաշվարկների նորարարական մեթոդների մշակմանը մոլեկուլային մոդելավորման համար, ունի բացառիկ **արդիականություն և գիտական նշանակություն:**

Ատենախոսությունը կազմված է հինգ գլուխներից և 71 գրականության ցանկից:

Առաջին գլուխը հանդիսանում է ներածություն, որտեղ հստակ ձևակերպված են աշխատանքի հիմնական նպատակները և խնդիրները: Հեղինակը համոզիչ կերպով հիմնավորում է բարձրորակ մոլեկուլային տվյալների գեներացման համար նոր մոտեցումների անհրաժեշտությունը, մատնանշելով գոյություն ունեցող տվյալների բազաների սահմանափակումները:

Երկրորդ գլուխը նվիրված է Smart Distributed Data Factory (SDDF) հարթակի ճարտարապետությանը և իրականացմանը, որը նախատեսված է լայնածավալ

քվանտային քիմիական հաշվարկներ իրականացնելու համար՝ օգտագործելով կամավոր հիմունքներով մարդկանց տրամադրած հաշվողական ռեսուրսները:

Երրորդ գլուխը ներկայացնում է ակտիվ ուսուցման համակարգը մոլեկուլային էներգիայի կանխատեսման և տվյալների գեներացման համար: Հեղինակը մանրամասն ուսումնասիրել է տարբեր գրաֆային նեյրոնային ցանցերի (GNN) ճարտարապետություններ և ընտրել է լավագույն 5-ը՝ ստեղծելով բազմազան մոդելների անսամբլ: Առանձնահատուկ ուշադրության է արժանի նմուշառման նոր մոտեցումների մշակումը՝

- SDDF-LOSSFN մոտեցումը, որը հիմնված է MACCS մատնահետքերի վրա աշխատող գծային ռեգրեսիոն մոդելների վրա, որոնք կանխատեսում են մոդելների անսամբլի սխալանքը:
- Նոր մոտեցում մոլեկուլային դինամիկայի հետազոտից կոնֆորմացիաների ընտրության համար՝ հիմնված ատոմային ուժերի փոխադարձ համաձայնության գնահատման վրա:

Կատարված փորձերը ցույց են տվել, որ ներկայացված մոտեցումները որակապես զգալիորեն գերազանցում են ինչպես պատահական այնպես էլ որոշ այլ ընտրության մեթոդներին:

Չորրորդ գլուխը նվիրված է FlashRMSD գործիքի մշակմանը սիմետրիայի ճշգրտմամբ արմատ միջին քառակուսային շեղում (SC-RMSD) մետրիկայի հաշվարկների համար: Ներկայացված երկփուլանի ալգորիթմը ապահովում է բացարձակ ճշգրտություն և 3-4 անգամ ավելի բարձր արագություն քան առկա գործիքները: Ներկայացվում է նաև FlashRMSD-M տարբերակը, որն իրականացնում է մինիմիզացված SC-RMSD հաշվարկ՝ օգտագործելով Kabsch ալգորիթմը:

FlashRMSD-ի կիրառությունները կառուցվածքային կենսահնֆորմատիկայում:

- Մոլեկուլային դոկինգի արդյունքների գնահատում և դիրքերի կլաստերիզացիա
- Մոլեկուլային դինամիկայի հետազոտի վերլուծություն
- Սպիտակուց-լիգանդ փոխազդեցությունների ուսումնասիրություն
- Կոնֆորմացիոն փոփոխությունների քանակական գնահատում

FlashRMSD-ի կիրառությունները քիմիական ինֆորմատիկայում:

- Վիրտուալ սկրինինգի արդյունքների ֆիլտրացիա և կլաստերիզացիա
- Քիմիական շտեմարանների բազմազանության վերլուծություն
- Ֆարմակոֆորային մոդելավորում և համեմատություն
- Կոնֆորմացիոն անսամբլների վերլուծություն

Հինգերորդ գլուխը ամփոփում է աշխատանքի հիմնական արդյունքները և եզրակացությունները:

Ամփոփելով գլուխներում ներկայացված նյութը՝ ատենախոսության հիմնական արդյունքները հետևյալն են.

- Մշակվել է SDDF հարթակը, որը կամավորական հաշվարկային նորարարական հարթակ է՝ նախատեսված մեքենայական ուսուցման ուղղորդմամբ մոլեկուլային տվյալների հավաքածուների ստեղծման համար:
- Մշակվել և ինտեգրվել է ակտիվ ուսուցման համակարգը SDDF-ի ներսում, որը համատեղում է նոր մոտեցումներ ինչպես մոլեկուլային այնպես էլ կոնֆորմացիոն նմուշառման համար:
- Ստեղծվել և հրապարակայնորեն հասանելի է դարձվել լայնածավալ, բարձրորակ քվանտային քիմիայի տվյալների հավաքածու (մոտ 2.75 միլիոն մոլեկուլային կոնֆորմացիաներ 2025 թվականի մայիսի դրությամբ) ENAMINE REAL տվյալների բազայից՝ wB97X/6-31G(d) DFT էներգիաներով:
- Մշակվել են GNN մոդելներ կոնֆորմացիոն էներգիայի ճշգրիտ կանխատեսման համար՝ օգտագործելով SDDF-ի միջոցով ստեղծված տվյալները, որոնք գերազանցում են առկա մի շարք մոդելների որոշ թեսթային հավաքածուների վրա:
- Մշակվել է FlashRMSD՝ նոր, բարձր արդյունավետությամբ մեթոդը և գործիքը՝ SC-RMSD մետրիկայի հաշվարկի համար, որը զգալիորեն գերազանցում է գոյություն ունեցող գործիքներին արագությամբ և հուսալիությամբ :
- Ստեղծվել և հրապարակվել է համապարփակ մոլեկուլային կառուցվածքների հավաքածու SC-RMSD հաշվարկող գործիքների գնահատման համար:

Ատենախոսությունն, ընդհանուր առմամբ, խիստ ամբողջական է և համապարփակ: Սակայն, ինչպես ցանկացած գիտական աշխատանք, այն նույնպես կարող է շարունակական կատարելագործումների առարկա դառնալ: Հաշվի առնելով աշխատանքի հսկայական ծավալը և բնույթը՝ առկա են մի քանի աննշան դիտողություններ, որոնք, այնուամենայնիվ, ամենևին չեն նվազեցնում ստացված արդյունքների կարևորությունն ու արժեքը: Մասնավորապես, նկատառումները վերաբերում են հետևյալ կետերին.

- **Մոդելների որակական գնահատականը ավելի բարձր կարգի տեսությամբ հաշվարկված տվյալների վրա:** Ցանկալի կլիներ տեսնել ներկայացված ճարտարապետություններով մոդելների որակական գնահատականը, թեկուզ փոքր չափի տվյալների հավաքածուի վրա, որը հաշվարկված կլիներ ավելի ճշգրիտ բազիսային բազմություններով (օրինակ՝ triple-zeta): Սա կարևոր է GNN մոդելների ներուժը և դրանց սահմանափակումները հասկանալու համար:
- **FlashRMSD գործիքի աշխատանքը մակրոմոլեկուլների վրա:** Հետաքրքիր կլիներ տեսնել FlashRMSD գործիքի արդյունավետությունն ու կիրառելիությունը մակրոմոլեկուլների (օրինակ՝ սպիտակուցներ, նուկլեինաթթուներ) վրա:

Հեղինակի կողմից հրապարակվել է 3 գիտական աշխատություն, որոնցից մեկը միջազգային հեղինակավոր ամսագրում: Սեղմագրի բովանդակությունը լիովին

համապատասխանում է ատենախոսության հիմնական դրույթներին: Հետազոտության թեման և արդյունքները լիովին համապատասխանում են Ե.13.05 «Մաթեմատիկական մոդելավորում, թվային մեթոդներ և ծրագրերի համալիրներ» մասնագիտությանը:

Վահագն Նորիկի Ալթունյանի ատենախոսությունը ավարտուն գիտական աշխատանք է, որն ունի կարևոր գիտական և կիրառական նշանակություն մոլեկուլային մոդելավորման և տվյալների գեներացիայի ոլորտում: Աշխատանքը համապատասխանում է թեկնածուական ատենախոսություններին ներկայացվող պահանջներին, և հեղինակը արժանի է Ե.13.05 «Մաթեմատիկական մոդելավորում, թվային մեթոդներ և ծրագրերի համալիրներ» մասնագիտությամբ տեխնիկական գիտությունների թեկնածուի գիտական աստիճանի շնորհմանը:

Պաշտոնական ընդդիմախոս՝

ՀՀ ԳԱԱ ԻԱՊԻ կենսաինֆորմատիկայի լաբորատորիայի
ղեկավար, ֆիզ.-մաթ. գիտ. դոկտոր, դոցենտ

Ա.Հ. Պողոսյանի ստորագրությունը
հաստատում եմ,

ՀՀ ԳԱԱ ԻԱՊԻ գիտ. քարտուղար



Հ. Պողոսյան

Սահակյան

20.06.2025