

Պաշտոնական ընդդիմախոսի կարծիք

Վահագն Նորիկի Աթոնյանի «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:

Ատենախոսության մեջ քննարկվում է քվանտային քիմիայի, մեքենայական ուսուցման և բաշխված հաշվարկների ոլորտների խաչմերուկում գտնվող կարևոր և արդիական խնդիր, ոլորտներ, որոնք մեծ թափով զարգանում են և կարևոր են հաշվողական մոլեկուլային գիտության ապագայի համար: Ատենախոսության կենտրոնական գիտական նպատակը՝ մոլեկուլային հատկությունների կանխատեսման և բարձրորակ քվանտային քիմիական տվյալների բազաների ստեղծման ամբողջական հաշվողական համակարգի մշակումը, ուղղակիորեն համապատասխանում է հաշվողական քիմիայի և նյութագիտության ժամանակակից կարիքներին:

Ատենախոսության առանցքում ընկած է մոլեկուլային համակարգերի և դրանց քվանտային քիմիական հատկությունների ուսումնասիրությունը, մասնավորապես՝ մոլեկուլային կառուցվածքի (կոնֆորմացիոն փոփոխականությունը հաշվի առած) և չափելի հատկությունների միջև կապը: Այս ուղղվածությունը և՛ հիմնարար է, և՛ հեռանկարային, քանի որ այն անդրադառնում է ֆիզիկական քիմիայի կարևոր մարտահրավերներից մեկին՝ մոլեկուլային կոնֆորմացիայից հաշվելի հատկությունների ճշգրիտ և արդյունավետ ստացմանը, մի խնդիր, որն ավանդաբար հիմնված է թանկարժեք քվանտային մեխանիկական սիմուլյացիաների վրա: Աշխատանքը անդրադառնում է այս մարտահրավերին՝ միավորելով մեքենայական ուսուցման մոտեցումները թեթև և դինամիկ բաշխված հաշվարկային համակարգի հետ (որը հիմնված է կամավորական համագործակցության վրա)՝ մեծածավալ մոլեկուլային տվյալների ստեղծումը հեշտացնելու համար:

Ատենախոսությունը բաղկացած է 5 գլուխներից և գրականության ցանկից:

Առաջին գլխում ներկայացվում է մոլեկուլային հատկությունների կանխատեսման ոլորտում առկա հիմնական մարտահրավերները՝ հատուկ շեշտադրում կատարելով բարձրորակ քվանտային քիմիական տվյալների համակարգված և արդյունավետ ստեղծման անհրաժեշտության վրա: Այն ուրվագծում է ատենախոսության հիմնական նպատակները և ներկայացնում այն միջդիսցիպլինար համատեքստը, որում իրականացվել է ուսումնասիրությունը:

Երկրորդ գլխում ներկայացվում է Smart Distributed Data Factory-ի (SDDF)՝ կամավորական հիմունքներով գործող հաշվարկային պլատֆորմի նախագծումը և իրականացումը, որը մշակվել է լայնածավալ քվանտային քիմիական հաշվարկները հեշտացնելու համար:

Սկզբում ակնարկ է կատարվում բաշխված հաշվարկային առկա լուծումներին, որից հետո մանրամասն ներկայացվում է SDDF-ի ճարտարապետությունը:

Երրորդ գլխում ներկայացվում է ատենախոսության հիմնական արդյունքի՝ լայնածավալ քվանտային քիմիական տվյալների բազայի ստեղծման գործընթացը: Այն ներկայացնում է ակտիվ ուսուցման համակարգը, որն օգտագործում է GNN (Graph Neural Network) մոդելների անսամբլ՝ հետագա, ծանր քվանտային քիմիական հաշվարկների համար բազմազան և տեղեկատվական բարձր արժեքով մոլեկուլային կոնֆորմացիաների ընտրություն կատարելու համար: Գլուխն ամփոփում է տվյալների ստեղծման ամբողջական գործընթացը՝ սկզբնական կառուցվածքների ստեղծումից մինչև ՄՈՒ-ի (Մեքենայական Ուսուցում) կողմից ուղղորդվող կատարելագործումը՝ ընդգծելով մեթոդաբանական նորարարությունները, որոնք հնարավորություն են տվել արդյունավետ և մասշտաբային ձևով ստեղծել բարձրորակ տվյալների բազա:

Չորրորդ գլուխը անդրադառնում է մոլեկուլային կոնֆորմացիաների համար սիմետրիայով ճշգրտված արմատ միջին քառակուսային շեղման (SC-RMSD) հաշվարկի բարդ խնդրին: Այն ներկայացնում է FlashRMSD-ն՝ բարձր արդյունավետությամբ ալգորիթ, որը նախատեսված է SC-RMSD-ի արդյունավետ և ճշգրիտ հաշվարկման համար, և մանրամասնում է դրա նախագծման սկզբունքներն ու տեխնիկական առանձնահատկությունները: Այս գլխի մեկ այլ ներդրում է SC-RMSD գնահատման համար հատուկ թեսթային տվյալների բազայի ստեղծումը, որը հնարավորություն է տալիս համակարգված համեմատություն կատարել SC-RMSD հաշվարկի գործիքների միջև:

Հինգերորդ գլուխը ամփոփում է աշխատանքը:

Հիմնական ներդրումները

- **SDDF հարթակ** - Կամավորների համակարգչային ռեսուրսներով քվանտային քիմիական (DFT) հաշվարկների իրականացման հնարավորություն: Այս թեթև համակարգը բաշխում է մեկական կոնֆորմացիաներից բաղկացած առաջադրանքները կամավորների համակարգիչների միջև՝ թույլ տալով լայնամասշտաբ տվյալների ստեղծում համայնքային մասնակցության միջոցով:
- **Մոլեկուլային տվյալների ստեղծման ակտիվ ուսուցում և լայնամասշտաբ քվանտային քիմիական տվյալների բազաներ** – Ակտիվ ուսուցման համակարգ, ներդրված SDDF-ի մեջ՝ տեղեկատվական բարձր արժեք ունեցող մոլեկուլային կոնֆորմացիաների արդյունավետ ընտրության համար: Համակարգը նպաստել է ատենախոսության կենտրոնական արդյունքներից մեկի ձեռքբերմանը՝ մոտ 2.75 միլիոն մոլեկուլային կոնֆորմացիոն էներգիաներից բաղկացած մեծ, բազմազան և բարձրորակ քվանտային քիմիական տվյալների բազայի ստեղծմանը և հրապարակմանը:
- **Սիմետրիայով ճշգրտված RMSD հաշվարկի արագագործ մեթոդ և գործիք** - FlashRMSD-ի նախագծում և մշակում՝ սիմետրիայով ճշգրտված RMSD (SC-RMSD) հաշվարկի բարձր արագագործությամբ, բաց հասանելությամբ գործիք, որը

նախագծված է սիմետրիկ մոլեկուլային կոնֆորմացիաների ճշգրիտ համեմատության համար: Գործիքի հետ մեկտեղ թողարկվել է թեսթային տվյալների բազա՝ SC-RMSD մեթոդների ստանդարտացված գնահատման և այս ոլորտում գործիքների ապագա զարգացմանը նպաստելու համար:

Ատենախոսության ամենաարժեքավոր ներդրումներից մեկը՝ լայնածավալ, բարձրորակ քվանտային քիմիական տվյալների բազայի մշակումն ու հրապարակային թողարկումը, զգալի ջանքերի արդյունք է տվյալների ստեղծման, մշակման և ենթակառուցվածքների առումով, և ակնկալվում է, որ այն հետաքրքրություն կառաջացնի գիտական հանրության շրջանում: Բացի այդ, աշխատանքի գործնական կիրառելիությունն ապացուցվում է նաև այն փաստով, որ արդյունքներից շատերը արդեն ներդրվել են որոշ կազմակերպություններում և գիտական խմբերում:

Սակայն, այս ներդրումների կողքին, ատենախոսությունը ունի նաև մի քանի մեթոդաբանական և հայեցակարգային թերություններ, որոնք արժանի են ուշադրության:

- **Տվյալների բազայի օգտակարության գնահատման բացակայություն:** Թեև ատենախոսությունը մանրամասնորեն նկարագրում է տվյալների բազայի նախագծումը և կառուցումը, այն չի տրամադրում դրա օգտակարության որևէ էմպիրիկ գնահատում մեքենայական ուսուցման մոդելների ուսուցման գործում: Չկան փորձեր, որոնք համեմատում են այս տվյալների բազայի վրա ուսուցված մոդելները՝ այլ հայտնի թեսթային բազաների վրա այլ մոդելների հետ:
- **FlashRMSD-ի դերի և կիրառման անբավարար մանրամասներ:** Չորրորդ գլխում ներկայացվում է սիմետրիայով ճշգրտված RMSD հաշվարկի ալգորիթ, սակայն դրա գործնական դերը տվյալների ստեղծման համակարգում անորոշ է: Չկա քննարկում, թե ինչպես է այս ալգորիթը ազդում տվյալների բազայի ֆիլտրման վրա, և չկա նաև դրա հետագա ազդեցության վերլուծություն ուսուցման կամ տվյալների որակի վրա: Մասնավորապես, ինչպիսի RMSD-ի շեմ է ընտրվում երկու կոնֆորմացիաները համարյա նույնը համարելու համար, և նշված արժեքի տարբեր ընտրությունները ինչպես են ազդում մոդելների որակի վրա:
- **Տեխնիկական նկարագրություններում բացթողումներ:** Ատենախոսության մեջ որոշ տեղերում բաց է թողնվում որոշ մեթոդաբանական մանրամասներ: Օրինակ, երկրորդ գլուխը հղում է կատարում «ԱԲ-ով ուժեղացված առաջադրանքների ընտրիչին», սակայն հետագայում՝ երրորդ գլխում, չի ներկայացնում թե կոնկրետ որ կոմպոնենտն է համապատասխանում վերոնշյալ ընտրիչին: Այս բացթողումները դժվարեցնում են աշխատանքի ընկալումը որպես մեկ ամբողջական համակարգ:

Ընդհանուր առմամբ, չնայած վերոնշյալ թերություններին, ատենախոսությունը ներկայացնում է զգալի տեխնիկական աշխատանք և ներառում է մի քանի **կարևոր ներդրումներ**, այդ թվում՝ ենթակառուցվածքների մշակում, լայնածավալ տվյալների բազա, ակտիվ ուսուցման համակարգ և նոր գործիքներ տվյալների հարստացման և գնահատման համար:

Հեղինակի կողմից, ատենախոսության թեմայով հրապարակվել է 3 գիտական աշխատանք, որոնցից մեկը միջազգային հեղինակավոր ամսագրում: 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՝

Պիկսարտ ՍՊԸ Ինժինեռական թիմի ղեկավար, տեխ.
գիտ. թեկնածու



Վ. Գ. Վարդանյան

Վ. Գ. Վարդանյանի ստորագրությունը հաստատում եմ՝

Պիկսարտ ՍՊԸ, ՄՈՎ ղեկավար

Վարսենիկ Վարդանյան

26.06.2025

