

분산 처리 기반의 사용자 입력 예측을 이용한 분자 모델링 시뮬레이션 도구의 상호작용 시스템

김보순^o 김지인

건국대학교 HCI연구실

{foxgap@dreamwiz.com^o, jnkm@konkuk.ac.kr}

An Interactive Molecular Modeling Using a Motion Path Prediction on a Distributed Processing System

Bosoon Kim^o Jee-In Kim

HCI Lab. Konkuk University

요 약

분자모델링 시뮬레이션 도구는 신 물질과 신약개발을 위한 가상 실험을 하는데 사용되는 중요한 도구이다. 이 도구는 분자 물질간의 결합을 통하여 에너지 계산을 하여 가장 낮은 에너지 준위를 보여주는 위치를 탐색한다. 에너지 결과를 빠르게 제공하면 사용자가 분자 물질을 결합하여 안정된 위치를 찾는데 많은 도움을 줄 수 있다. 본 연구에서는 에너지 계산을 고성능으로 처리 할 수 있는 분산처리시스템과 실시간 응답성 보장 스케줄링 알고리즘을 적용한 환경에 사용자 입력 예측 시스템을 추가하여 에너지 계산을 기존 시스템 보다 빠르게 요청하였으며, 에너지 계산 결과의 응답성을 향상 시켰다. 또한 사용자 입력 예측 시스템에서 발생하는 오차의 문제를 예측 시스템의 운영 방식을 통해 해결 방법을 제시하고 있다. 본 연구는 에너지 결과의 응답성 보장을 통하여 사용자가 좀 더 빠르게 안정된 결합 위치를 찾도록 도와준다.

1. 서 론

분자 모델링 시뮬레이션 도구는 신 물질과 신약개발에 활용되며, 이에 필요한 생화학적인 실험을 컴퓨터를 통하여 모의 실험하기 위해서 사용된다. 생화학 분야의 연구자들은 컴퓨터에 모델링된 분자모델을 조작하여 결합하고 이를 기준으로 분자의 에너지 계산 결과를 보고 이들이 안정되게 결합되는 위치를 탐색한다. 이렇게 안정된 위치를 찾는 작업을 Docking 작업이라고 하며, 이 작업은 Receptor라고 불리는 분자 모델과 Ligand라 불리는 분자 모델이 결합되는지 여부를 모의 실험을 통하여 확인하는 작업이다. 이들이 결합하여 최소의 에너지 준위를 보여주는 위치를 Active Site라고 한다.

사용자는 분자 모델링 시뮬레이션 도구를 사용하여 빠르게 에너지 계산 결과를 찾아 두 물질이 결합하여 안정화되는 Active Site를 찾아내려고 노력한다.[1,2,3] 이 과정에서, 사용자가 Receptor와 Ligand를 입력 도구(Input device)를 통해 조작하게 된다. 사용자의 조작에 대하여 실시간으로 시뮬레이션 계산이 수행되어야 하는데 시뮬레이션 계산량이 너무 많아서 물질이 이동되는 모든 점에 대하여 계산을 하고 실시간으로 응답하는 것은 거의 불가능하다. 따라서 본 논문에서는 분자 모델의 이동 경로 상에 있는 모든 점에 대하여 시뮬레이션 계산을 수행하는 대신에, 일부 이동점에 대하여는 계산을 수행하고 이를 제외한 이동점에 대하여는 실제 시뮬레이션 계산을 하는 대신에 물질의 이동 경로를 추적하여 예상하는 동작 경로 예측 시스템 (Motion Path Prediction)을 적용하여 실시간으로 경로 이동에 대응하는 방법을 개발하였다.

본 연구에서는 빠른 시뮬레이션 계산을 위하여 분산 처리 시스템을 구축하였고, 이를 위한 실시간 작업 반환 알고리즘을 개발하여 분자 모델링 작업을 효율적으로 수행할 수 있도록 하였다.

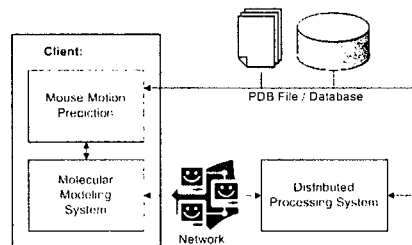
2. 관련 연구

RASMOL은 분자모델링 도구로 3차원 분자모델 구조를 보여준다. 범용적으로 사용되며, 빠른 3차원의 분자모델 시각 환경을 제공한다. 그러나 실시간 시뮬레이션 환경은 제공하지 않는다.

고성능의 연산을 제공하기 위한 다양한 방법이 존재하며, 컴퓨팅 자원을 인터넷에서 수집하여 사용하는 예도 존재한다. GRID Computing[4]은 이용하여 안정성, 고성능 그리고 신뢰성이 필요한 군사, 과학, 학술 그리고 산업 등에 많은 컴퓨팅 자원을 공급하는 것을 목적으로 하고 있다. 연산을 위한 목적, 자료의 보관과 공유 그리고 예측 및 평가 시스템으로 활용되고 있다. SETI(Search for Extra Terrestrial Intelligence)[5]는 외계 지적 생명체 탐사를 위한 프로젝트이다. 인터넷에 연결된 다수의 컴퓨터를 사용하여 외계지적생명체를 찾아내는 과학 실험이다.

3. 시스템

3.1 시스템 구조



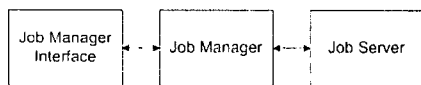
[그림1] MMWeb 시스템 구성

클라이언트는 분자모델링 시뮬레이션 도구이며 클라이언트는 에너지 계산을 고속으로 수행하기 위해 분산처리시스템과 연동한다. 클라이언트와 분산처리시스템 모두 분자모델링 데이터를 공급 받는다. 이러한 분자모델링의 데이터는 PDB(Protein Database)라는 형태로 규약된다.

본 연구에서는 MMWeb (Molecular Modeling for Web) [6]을 사용한 3차원 기반의 분자모델링 시뮬레이션 도구에 사용자의 동작 경로 예측 (Motion Path Prediction) 시스템을 적용하여 미리 사용자가 이동하는 분자모델의 위치를 기반으로 분산처리시스템을 이용하여 에너지 계산을 수행한다.

3.2 분산처리시스템

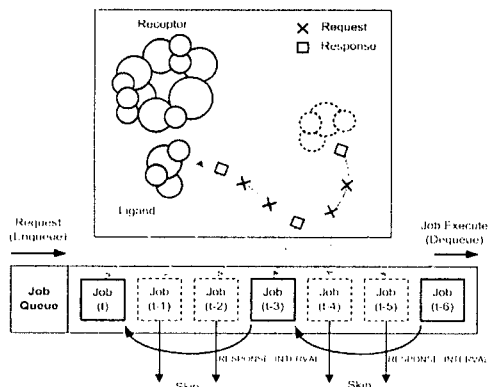
분자 모델링 시뮬레이션 도구에서 에너지 계산 요청을 고속으로 수행하기 위하여 분산처리시스템을 활용하였다.



[그림2] 분산처리시스템

그림 2에서 Job Manager Interface는 클라이언트가 접속하여 에너지 계산 작업을 요청하기 위한 게이트웨이(gateway) 역할을 한다. Job Manager Interface는 요청된 작업을 Job Manager에게 전달하고 Job Manager는 Job Server를 관리하고 요청된 작업을 Job Server에 배분하는 스케줄링을 담당한다.[6]

3.3 Job Skip Operation

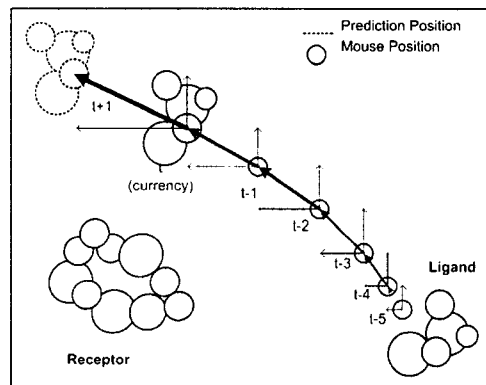


[그림3] Job Queue에서 Job Skip Operation

사용자가 분자모델링 시뮬레이션 도구와 상호작용하면서 Receptor와 Ligand를 조작할 때 일정 시간(Interval)마다 분산처리시스템에 에너지 계산을 요청하게 되며, Job Queue에는 이러한 클라이언트의 요청된 작업이 쌓이게 된다. 시간 순서로 쌓여진 모든 작업이 완료되기 위해서는 상당히 많은 시간을 필요로 하기 때문에 실시간으로 에너지 계산 결과를 클라이언트에게 제공할 수 없다. Job Skip Operation(그림3)은 응답 시간(RESPONSE_INTERVAL)을 기준으로 건너뛰면서(Skip) Job Queue에 쌓인 작업 가운데 시간상으로 고르게 일정한 수의 작업만을 선별 처리하여 에너지 계산 결과의 실시간 반환을 보장한다.[6]

3.4 사용자 입력 예측 시스템

에너지 계산 작업에는 Receptor와 Ligand의 후보물질에 대한 정보와 위치 정보가 사용된다. 분자모델링 시뮬레이션 도구의 사용자 인터페이스에서 사용자가 Receptor와 Ligand의 위치를 변경시킬 때 과거의 위치 벡터로 다음의 위치를 예측할 수 있다. 우리는 Mouse Motion Path Prediction을 사용하여 미리 후보물질의 다음 위치를 예측하여 에너지 계산 작업의 위치 정보로 활용하였다.



[그림4] 사용자 입력 예측

2차원에서는 2개의 축(x,y), 3차원에서는 3개의 축(x,y,z)으로 사용자의 입력에 대한 위치 정보를 각 성분 벡터로 분리한다. 과거의 시간 $t-1, t-2, \dots, t-n$ 에 대해서 위치를 Queue에 기록한다. 그리고 현재 시간 (t)에 대해서 미래의 시간 $t+\alpha$ 의 위치를 예측한다.

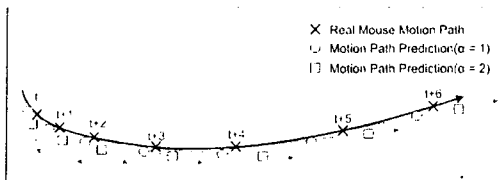
예측은 기존의 성분 벡터 정보의 벡터 합으로 얻을 수 있으며, 좀 더 정밀한 예측을 위해서 현재 시간 (t)에 가까워 질수록 가중치를 증가시켜 현재 시간에 가까운 위치정보가 다음 위치를 예측하는데 큰 비중을 차지하도록 한다. 이러한 이유는 사용자가 갑작스런 가속도의 변화에 대해 현재의 모멘트를 증가시키기 위한 방법이다.

사용자 입력 예측 시스템의 적용은 크게 사용자가 1) 일정 속도로 위치를 이동시킬 때와 2) 가속도를 가지고 이동시킬 때 그리고 3) 정지할 때 등으로 구분하여 예측의 깊이가 다르다.

일정 속도로 위치를 이동시킬 때는 동일한 벡터의 연속이므로 $t+\alpha$ 의 위치는 기존의 위치 변화와 동일하다. 가속도를 가지고 이동시킬 때 가속도가 증가하면 증가할수록 $t+\alpha$ 의 위치는 사용자가 움직이는 방향에 대해서 더 멀리 떨어진 위치를 예측한다. 위치 변화가 계속되는 가운데 정지를 하게 되면, 속도가 0이므로 $t+\alpha$ 의 위치는 시간이 갈수록 현재 위치에 수렴한다.

4. 실험

실험에 사용된 Receptor는 HIV-1(Human Immunodeficiency Virus)[8,9]을 사용하였으며 Ligand는 HIV-1의 재생(reproduction)과 관련된 물질을 Receptor와 Ligand의 분자량을 고정하였다. 분자모델링 시뮬레이션 도구의 사용자 인터페이스 상에 적용한 입력 예측 시스템(Mouse Motion Path Prediction)을 사용하여 $t+\alpha$ 의 예측과 실제 $t+\alpha$ 에서의 마우스 위치를 얻어 에너지 계산 결과를 비교하였다. α 는 1과 2로 하여 짧은 예측과 긴 예측으로 나누어 실험하였다.

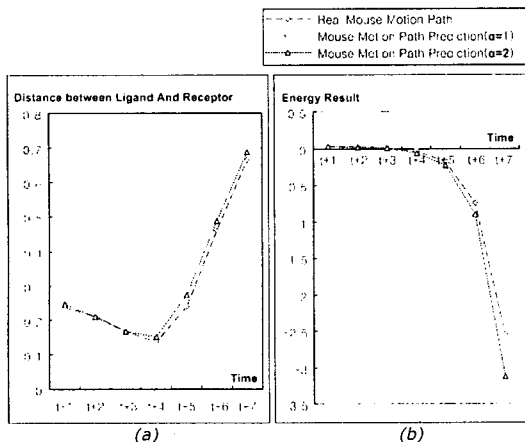


[그림5] 마우스 예측 위치와 실제 위치

Time	Ligand	Distance between Ligand And Receptor	Energy Result
	Real Mouse Motion Path		
t	(0.071, 0.357, 0.000)	0.236	0.034
t+1	(0.107, 0.386, 0.000)	0.211	0.025
t+2	(0.179, 0.414, 0.000)	0.166	0.005
t+3	(0.321, 0.436, 0.000)	0.138	-0.047
t+4	(0.501, 0.429, 0.000)	0.239	-0.181
t+5	(0.752, 0.392, 0.000)	0.461	-0.743
t+6	(0.964, 0.344, 0.000)	0.665	-2.549
Ligand Mouse Motion Path Prediction($\alpha=1$)			
t	(0.052, 0.344, 0.000)	0.252	0.038
t+1	(0.083, 0.382, 0.000)	0.232	0.029
t+2	(0.169, 0.424, 0.000)	0.180	0.006
t+3	(0.302, 0.441, 0.000)	0.141	-0.040
t+4	(0.488, 0.449, 0.000)	0.240	-0.173
t+5	(0.708, 0.411, 0.000)	0.423	-0.597
t+6	(0.947, 0.355, 0.000)	0.649	-2.378
Ligand Mouse Motion Path Prediction($\alpha=2$)			
t	(0.064, 0.370, 0.000)	0.246	0.033
t+1	(0.113, 0.394, 0.000)	0.209	0.022
t+2	(0.190, 0.428, 0.000)	0.168	0.005
t+3	(0.344, 0.447, 0.000)	0.153	-0.060
t+4	(0.532, 0.446, 0.000)	0.274	-0.225
t+5	(0.778, 0.403, 0.000)	0.488	-0.898
t+6	(0.988, 0.347, 0.000)	0.689	-3.118

[표1] 마우스 예측 위치와 실제 위치 좌표의 에너지 결과

그림 5는 사용자의 위치 이동에 따른 예측 경로와 실제 경로를 짧은 예측($\alpha=1$)과 긴 예측($\alpha=2$)으로 구분한 것이다. 표 1은 그림 5의 시간에 따른 위치 좌표와 위치 좌표에 대한 Receptor와 Ligand의 에너지 계산 결과를 산출한 것이다.



[그림6] (a)Ligand의 입력 예측과 실제 위치에 대한 Receptor와의 거리 (b)입력 예측과 실제 위치에 대한 에너지 계산 결과

그림 6(a)는 표1의 Ligand의 입력 예측과 실제 위치에 대한 Receptor와의 거리를 그래프로 나타낸 것이며 그림 6(b)는 표 1의 예측과 실제 위치에 대한 에너지 계산 결과를 그래프로 나타낸 것이다. 그림 6(a)와 그림 6(b)에서 Ligand와 Receptor의 거리를 기준으로 하였을 때 Ligand의 실제 위치상의 거리와 예측 위치($t+a$)상의 거리의 차이가 커지면 커질수록 에너지 계산의 오차도 증가 하는 것을 알 수 있다. 또한 이를 근거로 가속도가 크면 클수록 Ligand의 실제 위치와 예측 위치간의 오차가 심하게 발생한 다는 것도 알 수 있다. 그러나 에너지 계산 결과의 형태는 예측과 실제 위치 모두 비슷하게 나타난다.

에너지 계산 결과의 오차는 입력 속도가 줄어들거나 멈추면 실제 위치에 수렴하여 줄어든다.

5. 결론

입력 예측 시스템을 적용한 실시간 에너지 계산 결과의 보장은 예측 시간만큼 응답성이 보장된다는 점에서 의미가 있다. 한 걸음 더 나아가 예측 위치에 대한 에너지 계산 결과는 실제 위치의 속도가 줄어들거나 멈추게 됨에 따라서 오차는 줄어든다. 이 사실을 통해 응답성을 향상시키기 위한 입력 예측 시스템(Motion Path Prediction)을 적용한 경우의 오차 문제를 입력 예측 시스템 스스로가 보완 해준다.

이러한 입력 예측시스템의 장점을 활용하여 분자모델링 시뮬레이션 도구 상에서의 후보물질 Docking작업을 좀 더 빠르고 원활히 수행 할 수 있을 것으로 예상된다.

참고문헌

1. Ekachai Jenwitheesuk and Ram Samudrala, Improved prediction of HIV-1 protease-inhibitor binding energies by molecular dynamics simulations, BMC Structured Biology, 3:2, 2003.
2. James C.Philips, Gengbin Zheng, Sameer Kumar and Laxmikant V.Kale, NAMD:Biomolecular Simulation on Thousands of Processors, IEEE, pp36, 2002.
3. B.R.Brooks, R.E.Brucoleri, B.D.Olafson, D. Vid J. States, S. Swaminathan, and M.Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J.Comp. Chem., 4:187-217, 1983.
4. The Globus Project, <http://www.globus.org>.
5. The SETI Project, <http://www.seti.org/>
6. SungJun Park, Bosoon Kim, Jee-In Kim, "A Interactive Molecular Modeling System based on Web Service", FSKD 2005, LNCS Vol. 3613, pp1117-1126, 2005(SCIE)
7. Addison Chan, A Hybrid Motion Prediction Method for Caching and Prefetching in Distributed Virtual Environments, ACM VRST, pp135-142, 2001.
8. Earl Rutenber, Eric B.Fauman, Robert J.Keenan, Susan Fong, Paul S.Furth, Paul R.Ortiz de Montellano, Elaine Meng, Irwin D.Kuntz, Dianne L.DeCamp, Rafael Salto, Jason R.Rose, Charles S.Craik, and Robert M.Stroud, Structure of a Non-peptide Inhibitor Complexed with HIV-1 Protease, The Journal of Biological Chemistry, Vol. 268, No. 21, pp.15343-15346, 1993.
9. Junmei Wang, Paul Morin, Wei Wang, and Peter A. Kollman, Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA, Journal of American Chemical Society, Vol. 123, pp. 5221-5320, 2001.