안녕하세요. 저희는 적대적인 환경에서 Point-and-Click Behavior를 모델링하는 연구를 진행한 박진형, 심규철, 이현우입니다.

발표는 다음 순서로 진행하도록 하겠습니다.

연구 소개에 앞서 용어 정의부터 하면 Point-and-Click Behavior Modeling이란 인간의 Point-and-Click Task 수행과정을 실제와 유사하도록 수학적인 모델링을 하는 것을 의미하는데요.

저희는 특수하게 “적대적 환경”에서의 Point-and-Click Behavior Modeling 하는 것을 연구 주제로 삼았습니다.

저희 연구가 기존 연구들과 차별되는 점은 크게 2가지가 있는데요. 첫 번째로는 모든 기존 연구들은 적대적인 환경이 아니라 agent가 1명인 간단한 상황을 가정하고 모델링했습니다. 하지만, 실제 세계에서는 User가 Point-and-Click Task를 다른 유저들과 경쟁하는 경우가 많겠죠. 따라서 저희는 최초로 적대적인 상황을 고려하여 모델링을 진행했으며, 이 점이 큰 차별점을 지닙니다.

두 번째로 대부분의 기존 연구들과는 달리 저희 모델은 인간의 의사결정과정과 손목의 움직임 등 인간의 특성을 직접적으로 고려해서 모델링한다는 점에서도 차별점이 있습니다.

그럼 지금부터 저희가 연구를 진행한 방법에 대해 말씀드리겠습니다. 저희는 적대적 강화학습을 통해 두 개의 모델을 경쟁적으로 학습시켰는데요. 그 적대적 강화학습 환경을 어떻게 설계했는지 말씀드리겠습니다.

먼저, 적대적 강화학습 환경은 중간발표 때 말씀드린 것과 대부분 동일하기 때문에 시간 관계상 중요한 부분을 중심으로 설명드리겠습니다.

먼저, 저희는 매 episode마다 서로 동일한 target을 생성하는 2개의 환경을 생성한 후, 각 agent가 클릭을 시도하는데 걸린 시간과 클릭 성공여부를 독립적으로 측정했습니다.

그 후에는 이 결과를 기준으로 reward를 계산했는데요. 두 agent가 모두 클릭에 성공했다고 하더라도 그림처럼 agent 1이 agent 2보다 늦게 클릭에 성공한 경우에는, 일찍 성공한 순서대로 Reward에 차등을 줘서 부여했습니다. 이처럼 저희는 자기 자신과 opponent의 결과를 함께 고려해서 계산한 최종 reward를 기준으로 각 agent의 policy를 업데이트했습니다.

최종 Reward는 이 기준을 따라서 계산했는데요. 먼저 Reward는 Click의 성공여부에 따라 결정되는 Click Reward와 Agent가 Target을 클릭하기 위해 사용한 손목의 가속도, 즉 들인 노력을 정량화하는 Motor Effort로 구성되어 있습니다. Click Reward는 4가지로 나눴는데 클릭을 시도하기 전에는 0을, 클릭을 실패 했을 경우에는 -1을 부여했습니다. 단, 클릭을 성공한 경우에는 케이스를 분류했는데요. 두 agent가 모두 클릭에 성공한 경우, 먼저 성공한

agent에게는 14를, 늦게 성공한 agent에게는 9를 차등 부여했습니다.

Episode Termination의 기준은 두 Agent의 클릭 성공 여부의 관계없이 각 Agent에게 모두 Click 시도 기회를 1번씩 부여하고 그 기회가 모두 소진될 때 종료되도록 했습니다. 기존에는 한 agent가 클릭에 성공하면 다른 agent는 클릭 시도 여부에 관계없이 실패로 처리하여 그 즉시 episode를 종료했는데요. 그렇게 했더니 한 agent가 학습 초반에 여러 번 질 경우, 해당 agent는 아예 클릭을 하지 않고 그냥 가만히 있어서 모터 effort만을 줄이는 방향으로 잘못 수렴되기 때문에 기준을 바꾸게 되었습니다. 따라서, 각 agent가 기본적으로는 모두 click을 성공하는 것을 목표로 하되, 클릭 성공 속도에 따라 reward를 차등 부여하는 방식으로 변경했습니다.

그러면 지금부터는 적대적 강화학습 결과에 대해 설명드리겠습니다. 저희는 크게 2가지 소주제를 잡고 연구를 진행했는데요. 첫 번째로는 Human Factor 값이 서로 동등한 Agent 2개를 적대적으로 강화학습시키는 연구, 두 번째로는 Human Factor 값이 서로 다른 Agent 2개를 적대적으로 강화학습시키는 연구를 진행했습니다.

Human Factor는 agent가 action을 선택할 때 고려하는 여러 변수들 중 하나인데요, Human Factor의 종류와 개수는 지도교수님의 기존 연구와 동일하게 사용했습니다.

그래서 첫 번째 실험으로는, 이러한 Human Factor 값들을 기존 연구에서 사용한 값들과 동일한 값으로 고정시켜서 사용하되, 한 개의 agent가 아닌 2개의 동일한 agent를 적대적으로 강화학습시킴으로써 Optimal Policy가 어떻게 달라지는지 비교하는 연구를 진행했습니다.

먼저 human factor가 동일한 두 개의 agent를 적대적으로 강화학습 시킨 결과, 두 agent는 동일한 양상으로 Reward가 누적됬습니다.

또한, 두 agent의 클릭 실패율도 동일하게 1.0에서 0.4 수준으로 감소하면서 수렴했습니다.

그 후에는 이렇게 적대적으로 학습시킨 모델을 적대적으로 학습시키지 않았던 기존의 교수님의 모델과 비교해보았습니다. 먼저 적대적으로 학습시킨 저희의 모델은 사실 2개의 agent를 학습시킨 것이기 때문에 2개의 모델이 있는데요. 발표 시간상 자세히 설명은 못드리겠지만 검증 결과 이 두 개의 모델은 서로 거의 동일했습니다. 따라서 이 두 개의 모델 중 한 개의 모델을 선택해서 이를 적대적으로 학습시키지 않은 모델과 비교했습니다. 비교는 정성적, 정량적 방법으로 모두 진행했습니다.

정성적 방법으로는 먼저 두 agent가 각각 타겟을 클릭하는 과정을 시각화해서 비교해보았습니다. 왼쪽이 저희의 모델이고 오른쪽이 교수님의 모델입니다.

비교 결과, 타겟이 커서와 멀리 있는 초반부에는 두 agent의 이동 양상이 거의 동일했습니다. 하지만 타겟이 가까워지면 적대적 모델은 기존 모델보다 더 짧은 미래까지만을 예측하여 빨리 클릭을 시도하는 경향이 있었고, 반면 일반 모델은 더 먼 미래까지를 고려해서 타겟과 자신의

이동경로를 비슷하게 따라가면서 클릭을 시도하는 경향이 있었습니다.

정량적으로 비교한 결과도 비슷한 것을 의미했는데요. 클릭을 시도하는 데까지 걸리는 시간을 의미하는 trial completion time과 click failure rate를 1만개의 에피소드에 대해 비교한 결과, 저희의 적대적 모델이 항상 기존 모델에 비해 Trial Completion Time이 짧다는 결과를 보였습니다.

따라서 결론을 정리하면 적대적 모델과 일반 모델의 초기 policy는 비슷하지만, target에 가까워졌을 때 적대적 모델은 마치 상대방을 의식하여 더 짧은 미래까지만을 예측하여 빨리 클릭을 시도하는 경향을 보였고, 그에 따라 trial completion time이 짧아지는 policy로 최적화되었습니다.

두 번째로는 Human Factor가 서로 다른 두 개의 agent를 적대적으로 강화학습시킨 결과입니다.

저희는 각 agent의 human factor들을 다음과 같이 변경했는데요. 첫 번째 agent는 Decision-Making Skill을 높였고, 두 번째 agent는 motor execution 능력을 높이도록 조정했습니다.

이렇게 설정하고 적대적으로 강화학습시킨 결과, 두 agent 중 Decision-Making Skill이 높은 agent가 reward를 더 빠르게 누적하는 양상을 보였고

클릭 실패율 또한 Decision Making Skill이 높은 Agent가 더 낮은 값으로 수렴했습니다.

따라서 학습 경과를 보여주는 그래프만 보아도 인간의 두 factor들 중 어떤 factor가 더 큰 영향력을 미치는지 가늠할 수 있지만, 마찬가지로 각 agent들을 정성적, 정량적으로 비교해보았습니다.

이전처럼, 각 agent의 point and click 수행과정을 비교한 영상입니다.

각 agent의 커서 이동 경로를 비교해보면 decision-making skill agent는 타켓의 현재위치로 바로 직진하는 경향이 있었고 motor execution이 향상된 agent는 다른 agent에 비해 먼 미래까지 고려해서 Target에 대한 상대속도를 0으로 맞추려고 곡선으로 접근하는 양상을 보였습니다.

정량적으로 비교한 결과, Improved Decision-Making Skill Agent가 Trial Completion Time과 Click Failure Rate가 모두 작은 결과가 나와서, 클릭을 더 빠르게 시도하면서도 동시에 정확도가 더 높은 agent라는 것을 알 수 있었습니다. 따라서 이를 통해 Point and Click Task를 수행할 때에는 Motor Execution보다 Decision Making Skill이 더 중요하다는 것을 알 수 있었습니다.

그러면 마지막으로 연구의 최종 결론을 정리해서 말씀드리겠습니다. 먼저, 적대적으로 학습시킨 모델은 기존 모델에 비해 정성적, 정량적인 측면에서 Optimal policy가 달랐는데요. 적대적 모델이 Trial Completion Time을 더욱 낮추는 방향으로 최적화가 되었습니다. 따라서 실제 적대적 환경에서 경쟁할 때에는 복잡하게 생각하기보단 타겟의 짧은 미래까지만을 예측해서 바로 클릭을 시도하는 것이 최적의 전략이라는 것을 시사한다고 할 수 있습니다.

두 번째로 저희 연구는 Point-and-Click Task를 수행할 때에는 Decision-Making Skill이 Motor Execution보다 더 중요하다는 것을 강화학습을 통해 증명했습니다. 실제로 기존 연구 중에서, 프로게이머와 일반인은 Decision Making Skill 측면에서 두드러진 차이가 존재한다고 기술한 연구가 있는데요, 저희는 이 사실을 강화학습으로 다시 한번 증명할 수 있습니다.

마지막으로 이런 결론들을 활용한다면, 적대적인 환경에서 Point-and-Click Task를 수행하는 사람들에게 좋은 가이드라인이 될 수 있을 것이고, 그 외에도 인간과 Point-and-Click을 겨루는 Agent를 개발하는 데 활용될 수 있을 것이라고 생각합니다.

이상 발표 마치겠습니다. 감사합니다.