**< 캡스톤디자인 최종보고서 >**

프로젝트 명: 사용자와 교감하는 대화형 인공지능을 위한 백채널 예측 모델 개발

팀원(참여학생): 이재형(소프트웨어융합학과), 최진욱(소프트웨어융합학과)

**요 약**

MotionMatching이란 가상 캐릭터 애니메이션을 생성하는데 사용되고 있는 기술로써, Clavet와 Büttner에 의해 연구되어 발표되었습니다.[1] MotionMatching기술은 애니메이션 클립을 블렌드 트리(Blend Tree)나 상태 기계(State Machine)에 배열할 필요가 없이, 캐릭터의 각 상황에 맞는 애니메이션 클립을 대규모애니메이션 데이터베이스에서 찾아 매칭 시킨 후 시퀀스들 간의 블렌딩과 전환을 연속적으로 수행하여 보다 자연스럽고 유연한 애니메이션을 생성할 수 있습니다. 때문에 어플리케이션 실행 시간에서 주어진 상황이 계속적으로 바뀌는 게임 산업에서 많이 활용되고 있습니다.

본 보고서에서는 기존 MotionMatching에서 단점을 알아보고, 뉴럴 네트워크를 활용한 딥러닝 기술을 도입하여 발전된 Learned MotionMatching[2] 을 구현하는 것을 목표로 과제를 수행 했습니다. 또한 구현된 “Learned MotionMatching” 기술을 바탕으로 사실적인 애니메이션을 출력하면서 몰입도를 높인 “모션 액션 택배 배달” 게임 컨텐츠를 제작 하였습니다.

# 서론

## 연구배경

상호작용가능한 어플리케이션, 특히 콘솔게임이나 PC게임 등에서는 점점 기술이 발전함에 따라 이용자는 AAA게임과 같은 더 사실적이고 몰입감 있는 동적인 세계를 원하며, 그들의 요구의 맞추어서 동적인 세계 안에서 시각적으로 보이는 그래픽 기술들이 어떻게 하면 사실 적으로 표현 될지에 관해서 많은 연구와 개발을 통해 발전 되고 있습니다. 여기서 시각적으로 보이는 그래픽 기술은 비단 완성도가 높은 게임 배경이나, 많은 양의 파티클을 요구하는 화려한 이펙트 뿐만 아니라 캐릭터의 움직임이 얼마나 사실적인가 또한 포함하고 있습니다.

이미 많은 사람들이 사실적인 캐릭터 애니메이션에 관하여 고민을 해왔고, 다양한 기술이 고안이 되었으며, 그 중 최근에 가장 많이 쓰이는 기술이 “MotionMatching”이라는 기술입니다. 하지만 “MotionMatching”기술에는 여러 단점이 존재 합니다.

첫 번째 단점은 애니메이션 데이터의 규모입니다. 모션 매칭은 방대한 양의 사전 녹화된 애니메이션 데이터를 필요로 합니다. 데이터베이스에 저장되는 애니메이션 클립이 많아질수록 메모리 사용량도 증가합니다. 메모리 사용량은 데이터 양에 따라 선형적으로 증가하여, 데이터가 많아질수록 더 많은 메모리를 요구합니다. 이는 고품질 애니메이션을 위해 많은 데이터를 필요로 하는 현대의 AAA 게임에서 특히 문제가 될 수 있습니다.

두 번째 단점은 **실시간 검색 비용이 많이 듭니다.** 최근접 이웃 검색(nearest neighbor search)을 실시간으로 수행해야 하므로, 많은 애니메이션 데이터가 있을 경우 런타임 성능이 저하될 수 있습니다. 리얼 타임이 중요한 게임 어플리케이션에서는 치명적으로 다가올수 있습니다.

세 번째 단점은 데이터베이스 구축 비용입니다. 대규모 애니메이션 데이터베이스를 구축하고 유지하는 데 많은 시간과 노력이 필요합니다. 또한 “MotionMatching”에서는 애니메이션 품질에 크게 의존 됩니다.  
원본 데이터가 부자연스럽거나, 부정확하면 결과 애니메이션도 품질이 떨어 질 수 있습니다.  
 네 번째 단점은 일반 “MotionMatching”을 사용했을 때 캐릭터간의 상호작용 혹은 캐릭터와 오브젝트 사이의 복잡한 상호작용이나 환경과의 복잡한 상호작용을 처리하는 데 어려움이 있습니다. 이는 더 복잡한 알고리즘이나 추가적인 데이터 처리 기술을 필요로 할 수 있습니다.  
 이러한 단점들 때문에 모션 매칭은 특정 상황에서는 매우 유용하지만, 모든 상황에서 최적의 솔루션은 아닙니다. 따라서 신경망 기반의 생성 모델을 활용하여 고안된 “Learned MotionMatching”을 기술의 구현해보자 과제 목표를 정했으며, 구현된 기술로 이용자가 즐길수 있는 컨텐츠를 만들고자 했습니다.

## 연구 목표

본 연구는 기존 “MotionMatching” 기술에서 뉴럴 네트워크를 활용한 딥러닝을 도입하여 향상된 “LearnedMotionMatching”을 구현하고, 동적으로 애니메이션을 생성하여 게임속에서 자연스러운 캐릭터 모션을 출력합니다.

## 기대 효과

기존 “MotionMatching”의 긍정적인 측면과 상황에 맞는 사실적인 동적 애니메이션 출력을 유지하면서 애니메이션을 데이터에서 출력을 하는 것이 아닌 신경망 기반의 딥러닝을 통해 상황에 맞는 적합한 애니메이션이 출력됨으로 써 한정된 메모리 안에서 다양한 애니메이션을 실시간으로 생성하고 출력하여 메모리와 런타임 성능을 개선 할 수 있습니다.

# 배경 지식 및 관련 연구

“Learned Motion Matching”은 캐릭터의 상황에 맞는 애니메이션을 매칭하기 위해 특징 벡터 구조를 사용합니다. 이 특징 벡터의 각 요소는 캐릭터의 미래 궤적 위치를 포함하며, 이 미래 궤적 위치는 Spring-Damper 알고리즘을 통해 계산됩니다.

Spring-Damper 미래 궤적 예측 기법으로 계산된 캐릭터의 특징 벡터는 모션 매칭 단계에서 파라미터 값으로 사용되며, 학습에 필요한 애니메이션 특징 벡터 데이터셋인 “FeatureDataSet”을 구성하는 과정에서도 활용됩니다. “Learned Motion Matching”에서는 학습에 필요한 “FeatureDataSet”뿐만 아니라, 특징 데이터에 대응되는 애니메이션 데이터셋 “PoseDataSet”도 구성해야 합니다.

이 두 가지 특징 벡터 데이터셋들을 구성한 후, 다음 세 가지 신경망 기술의 조합으로 상황에 적합한 애니메이션을 동적으로 생성합니다. 모션 매칭에 필요한 세 가지 신경망 과정은 다음과 같습니다:

1. **Compressor&Decompressor**
2. **Stepper**
3. **Projector**

이 세 가지 과정을 통해 특징이 매칭이되고 애니메이션이 생성되면, 이전 애니메이션과 다음 애니메이션을 매끄럽게 이어주는 블렌딩 알고리즘으로써 Inertialization Blending 기법이 사용됩니다.

## Ubsoft “LAFAN1” Dataset & BVH file Construction

학습에 필요한 애니메이션 데이터 셋으로 Ubsoft에서 제공하는 “LAFAN1[3]” 데이터 셋을 사용하였습니다..

이 데이터셋은 모션캡쳐 기술로 제작 되었으며, 파일 형식으로 .BVH 포멧으로 구성됩니다.

BVH (Biovision Hierarchy) 포맷은 모션 캡처 데이터를 저장하기 위한 파일 포맷입니다. 구성 요소로는 캐릭터 관절에 관한 계층 구조와 애니메이션 데이터를 모두 포함하는 텍스트 기반의 파일 형식입니다. BVH 포맷은 다음과 같이 두 부분으로 나뉩니다. 계층 구조 (Hierarchy) 섹션과 모션 데이터 (Motion Data) 섹션입니다.

* 계층 구조

계층 구조 섹션은 캐릭터의 뼈대 (skeleton) 구조를 정의합니다. 이 섹션은 다음과 같은 요소로 구성됩니다:

**- ROOT**: 루트 노드는 뼈대의 최상위 부모 노드입니다. 캐릭터의 중심 (예: 골반)으로 간주됩니다.

**- JOINT**: 각 조인트는 뼈대의 각 부분을 나타냅니다. 루트 노드 아래에 여러 조인트가 존재할 수 있으며, 계층 구조로 연결됩니다.

**- OFFSET**: 각 조인트의 상대적 위치를 정의합니다. 이는 부모 조인트로부터의 상대적인 위치를 나타냅니다.

**- CHANNELS**: 각 조인트의 모션 데이터를 설명합니다. 일반적으로 각 조인트는 3개의 회전 축 (Xrotation, Yrotation, Zrotation)을 가집니다.

**- End Site**: 말단 조인트를 정의합니다. 더 이상 하위 조인트가 없는 조인트를 나타냅니다.

### ■ 모션 데이터 (Motion Data) 섹션

모션 데이터 섹션은 각 프레임에 대한 모션 정보를 포함합니다. 이 섹션은 다음과 같은 요소로 구성됩니다:

**- Frames**: 총 프레임 수를 나타냅니다.

**- Frame Time**: 각 프레임의 시간 간격을 나타냅니다.

**- 모션 데이터 값**: 각 프레임에 대한 모든 조인트의 채널 데이터를 포함합니다. 루트 노드부터 시작하여 계층 구조에 정의된 순서대로 각 조인트의 데이터를 나열합니다.

## Feature & Pose Vector Construction

# 캐릭터의 특징 벡터와 애니메이션의 특징 벡터 데이터 셋의 구조는 다음과 같습니다.

# x = {t^t, t^d, f^t, f!^t, h!^t } ∈ R^27 으로 구성되어있고,

# t^t ∈ R^6 : Local 캐릭터로부터 지상에 투영된 미래궤적 위치벡터들 입니다. 각 요소는 애니메이션 60Hz기준 20, 40, 60 프레임의 정보를 가지고 있습니다. 이후 설명할 Spring-Damper 알고리즘 기법을 통해 구하게 됩니다.

# t^d ∈ R^6 : Local 캐릭터로부터 미래 궤적의 정면 방향입니다. 또한 마찬가지로 각 요소는 애니메이션 60Hz기준 20, 40, 60 프레임의 정보를 가지고 있습니다. 이후 설명할 Spring-Damper 알고리즘 기법을 통해 구하게 됩니다.

# f^t ∈ R^6 : Local 캐릭터로부터 두 발의 위치 벡터입니다.

# f!^t ∈ R^6 : Local 캐릭터로부터 두 발의 속도 벡터입니다.

# h^t ∈ R^3 : Local 캐릭터로부터 엉덩이 속도입니다.

특징벡터 구조에서 주의해야 할 것이 있습니다. 특징들은 그 크기가 매우 다를 수 있으므로, 이를 정규화하는 것이 중요합니다. 여기서 우리는 데이터셋 내 각 특징(예: 왼쪽 발 위치)을 표준 편차로 스케일링합니다. 그런 다음 이 스케일링은 검색에서의 중요성을 조정하기 위해 사용자 가중치를 통해 추가로 조정될 수 있습니다. 타겟 논문에서는 사용자 가중치를 1로 두고 구현 했습니다.

다음으로, 우리는 애니메이션의 단일 프레임에 대한 모든 포즈 정보를 포함하는 포즈벡터 Y를 구성합니다. 포즈벡터 데이터 셋의 구조는 다음과 같습니다.

y = {y^t, y^r, y!^t, y!^r, r!^t, r!^r, o\*} 으로 구성되어 있고,

y^t, y^r : 각 관절의 Local Translation과 Rotations 입니다.

y!^t, y!^r : 각 관절의 Local Translation과 Rotational Velocity 입니다.

r!^t, r!^r : 캐릭터 Root Translation과 Rotational Velocity 입니다.

o\* : 다른 추가적인 정보를 담고 있습니다. 예를 들어 발 접촉 정보, 같은 씬의 다른 캐릭터의 위치나 궤적, 캐릭터의 일부 관절의 미래 위치와 같은 특징들을 담고 있습니다. 이러한 추가적인 특징 정보를 이용하여 다른 캐릭터와 상호작용을 할 수도 있으며, 사물 및 환경과 상호 작용하는 등 여러 상황에 적합한 애니메이션 생성이 가능해 집니다. 본 프로젝트에서는 상호작용 단계는 배제를 했으며, 캐릭터의 움직임에만 집중하여 구현 하였습니다.

## Spring-Damper (미래 궤적 예측 알고리즘)

Spring-Damper로 미래궤적을 예측하기 전에 게임에서 자주 쓰이는 보간기법인 Damper 먼저 알아야 합니다. 우리가 객체를 이동시킬 때, 우리는 기존위치에서 목표 위치까지 갑작스러운 이동 보다는 부드럽고 자연스러운 위치 이동을 원합니다. 이때 사용 하는 것이 Damper 기법 이라는 보간 방법 입니다.

이것은 이동 보간 뿐만 아니라 회전 보간에도 응용되어 사용 할 수 있습니다. 이동 사이의 보간을 하는 방식은 다음과 같습니다.

두 지점 A와 B가 있고, 이들의 좌표가 각각 (x1, y1)과 (x2, y2)라 하겠습니다. 두 점 사이의 t라는 비율 (0 ≤ t ≤ 1)에 대해 보간된 점 P의 좌표 (xp, yp)는 다음과 같이 계산됩니다:

xp=x1+t×(x2−x1)

yp=y1+t×(y2−y1)

여기서 t는 보간 비율로, t=0이면 점 P는 점 A에 위치하고, t=1이면 점 P는 점 B에 위치하게 됩니다. t가 0과 1 사이의 값일 경우, 점 P는 A와 B 사이의 어느 한 지점에 위치하게 됩니다. 하지만 이 해결책에는 문제가 있습니다: 게임의 프레임 속도(또는 시스템의 시간 간격)를 변경하면 댐퍼의 동작이 달라집니다. 더 구체적으로, 프레임 속도가 낮을 때 객체가 더 느리게 이동합니다. 이것을 해결하기위해 The Exact Damper[4]를 고안했지만, 이 알고리즘에서는 목표 위치가 빠르게 변할 경우, 즉 이전 방향과는 상관없이 즉시 반대 방향으로 움직이는 불연속성이 발생할 수 있습니다. 이는 시각적으로 급격한 움직임을 초래할 수 있습니다. 예를 들어, 물체가 한 방향으로 움직이고 있더라도 목표가 방향을 바꾸면 즉시 반대 방향으로 움직이게 됩니다. 이러한 여러 복합적인 문제를 해결해서 나온 선형보간 알고리즘이 Spring-Damper 입니다.   
 Exact Damper의 문제는 속도의 연속성이 없기 때문에 발생합니다. 즉, 이전 프레임에서 발생한 변화가 댐퍼의 동작에 전혀 반영되지 않고 항상 목표 방향으로 이동하려고 시도하기 때문에 발생하는 것입니다.

이를 해결하기 위해 Spring-Damper는 현재 속도에 부드럽게 영향을 미치는 방식으로 개선될 수 있습니다. 이를 위해 현재 속도에 목표 방향으로 향하는 속도를 스케일링된 stiffness(가소성) 파라미터로 추가하는 방법이 제안됩니다. 또한 목표 속도를 나타내는 q 변수를 도입하여 현재 속도가 이 목표 속도로 점진적으로 접근할 수 있도록 damping(감쇠) 파라미터로 조절하는 방법도 포함됩니다.

결국, Spring-Damper는 간단한 댐퍼의 움직임을 흉내내면서도 목표 방향으로의 부드러운 이동을 제공합니다. 이 방식은 두 개의 "힘"을 사용하여 해결될 수 있습니다 - 목표 위치로 당기는 힘과 목표 속도로 당기는 힘입니다. 이를 통해 이전의 댐퍼와 비교하여 속도의 연속성을 유지하면서도 부드러운 운동을 가능하게 합니다.

우리는 이 개념을 확장시켜 미래궤적을 예측할 수 있습니다. 예를 들어 캐릭터의 현재 위치가 기준 위치가 되고, 조이스틱의 값을 적용하여 목표 위치를 정하게 되면 기준 위치에서 목표 위치까지의 궤적을 감쇠와 보간 방식으로 구할 수 있었습니다. 이는 이동 뿐만 아니라 회전에서도 똑같이 적용 시켜서 프레임마다 특징벡터의 미래궤적 정보를 업데이트 시켰습니다.

## Quaternion Rotation

## 애니메이션에서 사용하는 회전 방식은 크게 Euler angles방식과 Quaternion 방식을 사용합니다. 두 방식에는 단점과 장점이 분명하여 필요한 상황에 따라 어떠한 방식을 사용할지 결정 해야 합니다. 쿼터니언(Quaternion) 회전은 3차원 공간에서 오일러 각(Euler angles)을 사용하는 것보다 효율적으로 회전을 표현하고 계산할 수 있는 방법입니다. 쿼터니언은 하이퍼 복소수(hypercomplex number)로, 벡터와 스칼라 부분으로 구성됩니다. 일반적으로 q = w + x\_i + y\_j + z\_k 로 표현됩니다. 여기서 w, x, y, z는 실수입니다. 쿼터니언(Quaternion) 회전을 사용하면 효율적인 계산 뿐만 아니라Euler Angles의 치명적인 단점인 짐벌락 현상을 해결 할 수 있습니다. 그렇기 때문에 쿼터니언(Quaternion)방식은 공간에서의 정밀한 회전 제어가 필요한 많은 응용 분야에서 쿼터니언을 선호하고 있습니다.

## 본 보고서에서 진행된 “Learned MotionMatching”에서의 학습단계에서 애니메이션데이터셋을 뉴럴 모델에 학습 시킬 때 Quaternion 방식을 사용하게 되고 또한 뉴럴 네트워크 단계를 걸쳐 나온 출력 값

## 또한 Quaternion 회전 값으로 출력하게 됩니다. 우리는 쿼터니언으로 출력된 회전값을 캐릭터에 적용할 때 Euler angles각 변환을 이용해서 적용 하였습니다.

## Compressor & Decompressor

# Decompressor는 원래의 자세 데이터 Y를 메모리에 저장하지 않고도, 특정 프레임 xi의 특징 벡터를 이용하여 해당 자세 yi를 효과적으로 생성하고 관리하는 기술입니다. 이를 가능하게 하기 위해 우리는 "Compressor"라고 불리는 네트워크를 사용하여 자세 yi를 저차원 잠재 공간 표현 zi로 매핑합니다. 이 저차원 잠재 공간 표현 zi는 특징 벡터 xi와 연결 되어 "Decompressor"에 입력으로 제공됩니다. “Decompressor”는 이 입력을 기반으로 원래의 자세 yi를 재구성합니다. 이 과정에서 우리는 특징 벡터 x에 부족한 추가 정보가 무엇인지를 학습하고, 이 정보를 잠재 변수 zi에 인코딩하여 보다 정확하고 완전한 자세 재구성을 목표로 합니다. 이 과정은 AutoEncoder와 유사한 구조로 이루어져 있습니다.

# “Decompressor”의 핵심 요소 중 하나는 사용하는 손실 함수입니다. 단순한 평균 제곱 오차 손실은 움직임이 떨림이나 낮은 품질로 나타날 수 있기 때문에 이를 대체하기 위해 시각적으로 인식되는 오차를 최소화하는 손실 함수를 설계합니다. 이 손실 함수에는 Forward Kinematics를 사용하여 캐릭터 공간에서의 오차를 측정하고, 자세가 시간에 따라 부드럽게 변하도록 하는 속도 기반 손실도 포함됩니다.

# 훈련 절차는 주어진 애니메이션 데이터베이스 Y와 매칭 데이터베이스 X로부터 두 개의 프레임을 사용하여 진행됩니다. “Compressor” C를 사용하여 잠재 변수 Z를 찾고, 이를 “Decompressor” D를 사용하여 원래의 자세를 재구성합니다. 각 샘플링(두 프레임의 쌍)에 대해 이 절차를 설명하며, 미니 배치의 각 요소에 적용하여 네트워크 매개 변수 θC와 θD를 업데이트할 때 결과를 평균화합니다. 연산자 ⊖는 두 자세 사이의 차이를 계산하는 데 사용되며, 회전 차이에 대해서는 회전 매트릭스를 사용하여 계산합니다. 속도 손실에서는 자세의 속도 차이를 계산하지 않습니다.

# 가중치 w∗는 모든 자세 기반 손실에 대해 거의 동일한 가중치를 부여하고, 정규화 손실에는 작은 가중치를 부여하여 설정됩니다. “Compressor”에게 로컬 및 캐릭터 공간 입력을 모두 제공함으로써 정확도를 높일 수 있었으며, 이는 “Compressor”가 특징이 유용하다고 판단하면 직접적으로 잠재 공간으로 복사할 수 있음을 의미합니다. 훈련이 완료된 후, “Decompressor”는 다른 네트워크 없이도 이미 유용하게 사용될 수 있습니다. z의 차원을 작게 만들어 각 프레임 i에 대해 잠재 변수 zi를 계산하고, Y 대신 Z = [z₀, z₁, ..., zn⁻₁]를 저장함으로써 중요한 메모리 절약을 달성할 수 있습니다. 이는 Motion Matching 알고리즘의 동작에는 전혀 영향을 미치지 않습니다.

# 마지막으로, “Decompressor”가 Y를 메모리에 저장할 필요성을 없애지만, 많은 매칭 특징이나 잠재 변수를 사용할 경우 X와 Z를 여전히 상당히 크게 저장해야 할 수 있습니다. 본 구현 과정에서 참고한 Compressor&Decompressor에 대한 pseudocode는 다음과 같습니다.

## 텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진 자동 생성된 설명

## < DANIEL HOLDEN, Ubisoft La Forge, Ubisoft, Learned Motion Matching[5] >

## Stepper

## 초기에 매칭된 잠재 특성 벡터 x\_i, z\_i 가 주어진 상황에서 우리는 다음 프레임의 애니메이션을 정보를 얻기위해 연속된 애니메이션 벡터에 인덱스 값을 증가시켜서 검색해야 합니다. “Learned MotionMatching”이 아닌 기존 “MotionMatching”에서는 이과정이 어렵지 않습니다. 단순히 인덱스 “I”를 증가 시키고 X 또는 Z의 새로운 행을 참조하면 됩니다. 하지만 “Learned MotionMatching”에서는 X와 Z를 메모리에 저장하지 않고 다음 애니메이션 행을 찾는게 목적이기 때문에 이 과정은 비교적 단순 하지 않습니다. 우리는 이 과정을 Stepper라는 신경망 기반 학습 모델을 통해 진행하려고 합니다. 이는 특정 프레임 x\_i, z\_i ​에서 입력으로 받아들이고, 다음 프레임에서의 특성 벡터 x\_i+1, z\_i+1​을 생성할 수 있는 델타를 출력하는 네트워크입니다.

# 여기서 우리는 자기 회귀적 방식으로 네트워크를 훈련시키며, 길이 S의 짧은 창 내에서 특성 벡터 X와 잠재 변수 Z를 반복적으로 예측하여 다음 프레임에서 그들을 피드합니다. 훈련 샘플 하나에 대해 절차를 설명하였지만, 미니배치의 각 요소에 적용하고 θS를 업데이트할 때 결과를 평균화합니다. 가중치 w∗ 는 모든 손실에 거의 동일한 가중치를 부여하기 위해 설정됩니다. 훈련을 마치면, Stepper는 파이프라인의 스텝핑 부분을 대체하는 데 사용될 수 있으며, X나 Z에 의존하지 않고 일치 및 잠재 특성 벡터의 스트림을 생성할 수 있습니다. 압축기를 사용하여 초기 시작 상태를 찾을 수 있지만, 가장 가까운 이웃 검색에는 X와 Z를 메모리에 유지해야 합니다. 본 구현 과정에서 참고한 Stepper에 대한 pseudocode는 다음과 같습니다.

## 텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진 자동 생성된 설명

## < DANIEL HOLDEN, Ubisoft La Forge, Ubisoft, Learned Motion Matching[5] >

## Projector

## 위 챕터에서의 Stepper 라는 긴경망 모델로 시간에 따라 특징 벡터를 진행 시킬 수 있었습니다. 하지만 Stepper만을 이용해서 특징데이터셋인 X에 대해서 메모리를 독립 시킬 수 없었습니다. 왜냐하면 쿼리 벡터 x^에 수행되는 매칭 검색 과정에서는 여전히 X 벡터가 사용되야 하기 때문입니다. 이것을 해결하기 위해 타겟 논문에서는 Projector라는 신경망 모델을 사용합니다. Projector의 목표에 대하여 설명 하면 프로젝터(Projector)는 주어진 입력 벡터에 대해 매칭 데이터베이스에서 가장 가까운 이웃을 찾아 그에 해당하는 특징 벡터와 잠재 변수를 출력하는 네트워크입니다. 이 네트워크는 일반적으로 최근접 이웃 검색을 대체하고, 학습된 데이터를 기반으로 쿼리 벡터와 가장 유사한 벡터를 찾아주는 기능을 하게 됩니다.

## 과정은 이렇습니다. 매칭 데이터베이스의 특징 벡터 x를 제공하면, 우리는 노이즈 크기 nσ를 샘플링하여 이를 가우스 노이즈 벡터 n에 스케일링을 합니다. 이를 x에 추가하여 xˆ을 생성하고, 가장 가까운 이웃 k∗를 찾습니다. 그런 다음 프로젝터는 해당 특징 벡터와 잠재 변수 xk∗ 및 zk∗를 출력하도록 훈련됩니다 훈련 샘플에 대해 여기서는 단일 절차를 제시했지만, 미니배치의 각 요소에 적용하고 θP를 업데이트할 때 결과를 평균화합니다. 다양한 노이즈 크기를 샘플링함으로써 프로젝터를 다양한 크기의 교란에 강화시킵니다. 가중치 w∗는 모든 손실에 대해 대략적으로 동일한 가중치를 제공하도록 설정됩니다. .

## 최종적으로 프로젝터가 훈련되면, 학습된 모션 매칭 파이프라인이 완성됩니다. 최근접 이웃 검색 대신, 매 N프레임마다 사용자 쿼리 xˆ를 프로젝터 P를 통해 전달합니다. 그런 다음 각 프레임마다, 스테퍼 S를 사용하여 찾은 매칭 및 잠재 특징 벡터를 전진시키고, 디코딩하여 디컴프레서 D를 사용하여 포즈를 생성합니다.

## 이제 Project를 사용함으로써 메모리에 X와 Z를 저장할 필요가 없도록 구성 할 수 있습니다.

## 본 구현 과정에서 참고한 Projector에 대한 pseudocode는 다음과 같습니다.

## 텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진 자동 생성된 설명

## < DANIEL HOLDEN, Ubisoft La Forge, Ubisoft, Learned Motion Matching[5] >

## inertialization

Inertialization[6] 이란 David Bollo가 2016년 GDC 2018에서 'Inertialization: High-Performance Animation Transitions in Gears'라는 주제로 발표한 캐릭터 애니메이션 Blending 기법 입니다. Blending 기법이란 기본적으로 두 가지 다른 애니메이션 상태 사이의 자연스러운 전환을 가능하게 합니다. 예를 들어, 캐릭터가 걷는 상태에서 달리는 상태로 전환될 때, 이 과정이 자연스럽고 부드럽게 이루어지게 됩니다. 이는 게임에서 플레이어 경험을 향상시키는 데 중요한 역할을 합니다.

기존의 Blend 기법은 보통 애니메이션 상태 사이의 선형 보간(Linear Interpolation)을 사용하여 전환합니다. 이는 간단하고 빠르지만, 전환 과정에서 갑작스러운 변화가 발생하여 플레이어 경험을 낮추기도 합니다. 하지만 Bavid Bollo가 고안한 'Inertialization Blending 기법은 물리적 관성을 모델링하여 애니메이션 전환을 부드럽게 만듭니다. 예를 들어, 캐릭터가 움직임을 멈출 때도 일정한 관성을 고려하여 부드럽게 변화하도록 만듭니다. 이는 보다 자연스러운 애니메이션 전환을 가능하게 합니다. 또한 이 기술의 핵심 장점 중 하나는 높은 성능을 제공한다는 점입니다. 게임에서 매우 많은 양의 애니메이션 전환이 발생 하여 성능 문제를 일으킬수 있는데, 'Inertialization' 기술은 이를 최대한 효율적으로 처리하는 알고리즘으로 게임의 부드러운 플레이를 유지할 수 있습니다.   
스크린샷, 3D 모델링이(가) 표시된 사진

자동 생성된 설명

< David Bollo. 2016. Inertialization: High-Performance Animation Transitions in ’Gears of War’. In Proc. of GDC 2018 [6]>

위 참조 그림 처럼 팔이 위에서 아래로 내려올 때 기존 Blending 방식과 Inertialization은 궤적 차이를 보입니다. Inertialization Blending 기법이 조금 더 자연스럽게 팔이 떨어지는 것을 볼 수 있습니다.

## 기존 연구 문제점 및 해결 방안

본 프로젝트에서는 베이스라인 모델인 BPM\_MT의 문제점을 파악하고, 이를 해결하기 위한 방안을 모색해 보았다. 첫 번째 문제는 실세계 발화 데이터에서 백채널 카테고리별 샘플 수의 불균형이다. 이러한 불균형은 모델의 학습 과정에서 특정 카테고리에 편향되게 만들 수 있으며, 이는 실제 환경에서의 모델 성능 저하로 이어질 수 있다. BPM\_MT는 이러한 불균형을 효과적으로 다루는 메커니즘이 부족하다. 따라서 우리는 데이터 불균형을 해결하기 위한 방법을 모색해야 한다.

두 번째 문제는 오디오 처리에 있어서 BPM\_MT가 LSTM을 사용한다는 것이다. LSTM은 상대적으로 오래된 모델으로, 최근에는 더욱 발전된 오디오 처리 기술이 존재한다. LSTM보다 더 효율적이고 정확한 오디오 처리 방법을 사용하는 모델을 사용 하는 것을 고려해야 한다. 해결 방안으로 첫 번째 문제점인 데이터 불균형을 해결하기 위해 우리는 다음과 같은 접근 방식을 채택하였다.

* 데이터 불균형 해결

손실 함수 개선: 모델이 더욱 효과적으로 다양한 데이터를 학습할 수 있도록 손실 함수를 조정한다. 이를 통해 모델이 소수의 샘플에 대해서도 더 잘 학습할 수 있게 한다.

* 데이터셋 입력 방식 개선

데이터 입력 방식을 개선하여 모델이 더 균형 잡힌 데이터를 받도록 한다. 이는 모델의 일반화 능력을 향상시키는 데에 중요하다.

* 오버샘플링

실제 입력 데이터 수를 증가시키기 위해 오버샘플링을 진행한다. 이는 특히 소수의 샘플이 존재하는 카테고리에 도움이 된다.

* 오디오 모델 교체

Wav2Vec 2.0[7]과 HuBert[8] 모델 적용: LSTM을 대신하여 최신 오디오 처리 모델인 Wav2Vec 2.0과 HuBert를 적용한다. 이 모델들은 LSTM보다 오디오 데이터의 특성을 더 잘 포착하고, 결과적으로 더 정확한 백채널 예측을 가능하게 할 것이라 예상된다

# 추진 내용

## 팀 구성 및 역할

|  |  |  |
| --- | --- | --- |
| 이름 | 역할 | 수행업무 |
| 이재형 | 팀장 |  |
| 최진욱 | 팀원 |  |

## 전체 시스템 구성

…….

## 사용자 시나리오

---

## 핵심 기능 개발

1. LSTM의 대체

기존의 오래된 LSTM 모델 대신, BPM\_MT에 최적화된 Wav2Vec 2.0과 HuBert 모델을 도입함으로써 성능을 향상시키려고 한다. LSTM은 그 구조상 긴 시퀀스 데이터에서의 의존성을 학습하기에 유용하지만, 연산 부하가 크고 최근에 개발된 모델들에 비해 성능이 상대적으로 떨어진다. Wav2Vec 2.0과 HuBert 모델을 BPM\_MT에 적용함으로써, 더 빠른 학습 속도와 높은 정확도를 달성할 수 있을 것으로 기대한다. 이러한 Transformer 모델들은 self-attention 메커니즘을 통해 시퀀스 내의 각 요소들 사이의 관계를 더 효과적으로 학습할 수 있어, 최신 모델의 도입은 BPM\_MT의 성능을 크게 향상시킬 수 있을 것으로 예상된다.

…..

## …..

## 이슈 및 대응

……………

# 결과

## GPT를 통한 Text Augmentation 진행 시 SOTA

------------

## GPT의 결과

-------

--------------

# 결론

본 연구는 Focal Loss와 Transformer 기반 모델(Wav2Vec2-Large-XLSR- Korean과 HuBERT)을 활용한 백채널 카테고리 예측에 대한 새로운 접근법을 제안하였다. 기존의 CE Loss 대신 Focal Loss를 적용함으로써, 클래스 불균형이 심한 데이터에서 발생하는 오분류 문제를 완화시켰다. 이는 Focal Loss가 각 클래스에 대한 모델의 예측 오류에 다른 가중치를 부여함으로써 보다 균형 잡힌 학습을 유도하기 때문이다. 더욱이, LSTM 대신 Transformer 기반 모델을 채택함으로써, LSTM이 가진 시퀀스 데이터의 장기 의존성 학습에 대한 한계를 극복하고, self- attention 메커니즘을 통해 모델의 성능을 향상시켰다. 이러한 모델 구조의 변화는 특히 오디오 데이터 처리에 있어서 중요한 의미를 가지며, 실험 결과에서도 Wav2Vec2-Large와 Focal Loss의 조합이 가장 우수한 성능을 나타내었다.

본 연구의 중요한 성과 중 하나는 GPT를 이용한 Text Augmentation이었다. 이 방법을 통해 소수 클래스의 데이터를 인위적으로 확장함으로써, 클래스 불균형 문제를 효과적으로 완화시키고, 모델의 일반화 능력을 크게 향상시켰다. GPT를 사용한 Text Augmentation은 EDA 기법보다 더 효과적인 Text Augmentation임을 시사한다. GPT는 새로운 문맥과 다양한 표현을 지닌 데이터를 생성함으로써, 모델이 실세계의 다양한 상황에 대응할 수 있도록 훈련시킨다. 이는 백채널 카테고리 예측의 정확도를 현저히 향상시켜, 자연스러운 의사소통 분석에 큰 도움이 될 것으로 기대된다.

이러한 결과는 오디오 모델과 손실 함수의 선택이 백채널 예측의 정확성에 결정적인 영향을 미친다는 것을 보여주며, 특히 GPT를 이용한 Text Augmentation의 효과는 이 분야에서의 중요한 발전으로 평가될 수 있다. 그러나 실험에서는 LSTM과 CE Loss가 특정 조건에서 여전히 우수한 성능을 보였으며, 이는 손실 함수와 모델 구조의 선택이 단순히 클래스 불균형 문제를 해결하는 것을 넘어 다양한 요소를 고려해야 함을 나타낸다. 향후 연구에서는 loss 함수를 더욱 개선시키고, ELECTRA와 같은 더 진보된 언어 모델을 탐색함으로써, 백채널 카테고리 예측의 정확도를 더욱 향상시킬 방안을 모색할 예정이다. 이는 백채널 모델의 발전에 크게 기여할 것으로 기대된다.

참고문헌

[1] Michael Büttner and Simon Clavet. 2015. Motion Matching - The Road to Next Gen Animation. In Proc. of Nucl.ai 2015. https://www.youtube.com/watch?v= z\_wpgHFSWss&t=658s

# [2] DANIEL HOLDEN, Ubisoft La Forge, Ubisoft, Canada SIGGRAPH 2020 – learned Motion Matching <https://youtu.be/16CHDQK4W5k?si=CfZw2QgU5GXvKqGY>

# [3]Ubsoft-“LAFAN1”Dataset. <https://github.com/ubisoft/ubisoft-laforge-animation-dataset>

# [4] The Exact Damper. <https://theorangeduck.com/page/spring-roll-call#exactdamper>

# [5] DANIEL HOLDEN, Ubisoft La Forge, Ubisoft, Canada. <https://theorangeduck.com/media/uploads/other_stuff/Learned_Motion_Matching.pdf>

# [6] David Bollo. 2016. Inertialization: High-Performance Animation Transitions in ’Gears of War’. In Proc. of GDC 2018 https://youtu.be/BYyv4KTegJI?si=GDP1qliaLHo5fBDH