

World-First Developmental Curriculum Learning and the Emergence of Axis Structure

월드먼저 발달 커리큘럼 학습과 창발하는 축 구조

한혁진 (bokkamsun@gmail.com) · ORCID 0009-0007-4132-1707

반야 연구 시리즈 제4편

본 문서 CC BY 4.0 · 코드 Apache-2.0

초록

사람은 언어를 배우기 전에 감각과 공간과 논리의 감각을 먼저 발달시킨다. 본고는 그 순서를 모방한 발달 커리큘럼(developmental curriculum, 쉬운 단계에서 어려운 단계로 이어지는 학습 순서)으로 언어모델을 학습한다. 언어 이전에 크기, 거리, 온도 같은 감각 축(sensory axis, 한 성질을 재는 눈금)을 먼저 형성하고, 그 위에 언어를 얹는다. 핵심 물음은 온톨로지(ontology, 개념들이 자기 자리를 갖는 구조도)를 직접 짜 넣지 않아도 학습 순서만으로 필터와 임베딩이 스스로 온톨로지적 축 구조를 형성하는가이다. 본고는 이를 다음과 같이 측정했다. 첫째, 감각 개념이 무작위의 5배에서 27배로 축에 자리한다. 둘째, 순서가 있는 축(거리, 온도)은 순서로 정렬되고(순서 정합 최대 67퍼센트), 순서가 없는 범주 축(색, 맛)은 순서 없이 형성되어 둘이 명확히 갈린다. 셋째, 관련 개념끼리 무리를 이룬다(공간 개념끼리, 감각 개념끼리). 넷째, 채널 필터가 얇은 층 큰 붓에서 깊은 층 가는 펜으로 필터 사다리를 이루며 죽은 필터가 하나도 없다. 즉 학습 순서 자체가 필터를 온톨로지로 만든다. 이 판독은 학습을 건드리지 않는 비침습(non-invasive, 몸을 안 열고 재는) 프로브로 얻는다.

측정

이 글의 모든 정량 수치는 집필 환경에서 측정한 실측값이다(발달 단계별 열린 모델, 환경은 부록 B). 이론적 배경은 별도 공리 문헌 [1]에 있으며, 본고의 주장은 그 공리가 아니라 측정에서 나온다.

목차

1. 서론	5
2. 관련 연구	5
3. 방법	6
4. 결과	6
4.1 감각 축이 무작위의 5배에서 27배로 형성된다	6
4.2 순서 축은 순서로, 범주 축은 순서 없이 형성된다	7
4.3 관련 개념끼리 무리를 이룬다	8
4.4 채널 필터가 필터 사다리로 자기조직하며 죽은 필터가 없다	8
5. 한계와 다음 측정 과제	9
6. 결론	9
7. 후속 논문 예고	9
참고문헌	10

부록 A. 발달 단계	11
부록 B. 측정 환경	11
부록 C. 재현 안내	11

용어 범례

용어를 아래 표에 정의한다. 같은 개념에는 같은 이름만 사용하며, 일곱 편의 본문과 그림이 모두 이 표를 따른다. 코드 기준은 동봉 코드에서 그 개념을 구현한 식별자이고, 일반 용어는 학계에서 통용되는 가장 가까운 이름이다. 다를 때만 칸을 채웠다.

용어 범례 1. 엔진과 신용

용어	뜻	코드 기준	일반 용어
정확한 전치	각 연산의 수반을 손으로 유도해 닫힌형으로 구현한 역방향 계산. 이 엔진과 신용 사슬의 고유명	banya_core.py	수동 역전파
수반	순전파의 짝 연산. 기울기를 정확히 거꾸로 되돌려 흘린다	bwd 커널들	adjoint
신용	어떤 지점의 은닉에 대한 손실 기울기. 파라미터가 결과에 진 책임	델타(δ)	그레이디언트
신용 사슬	캐시를 역순으로 되짚으며 수반을 곱해 신용을 입력까지 되돌리는 골격	backward	오차역전파
수반 정합	수반 식이 유한차분과 같은 값을 내는지 확인하는 검사	각 편 프로브	gradient check
순환행렬 채널 혼합	채널 혼합을 순환행렬로 두고 rfft 성분곱으로 계산하는 연산		순환 합성곱
상대 거리 혼합	거리에만 의존하는 공유 계수로 자리들을 섞는 혼합	m_mat_w_lag	Toeplitz 혼합
시간혼합	다른 자리의 정보가 현재 자리로 섞여 드는 경로의 총칭		토큰 혼합
혼합 헤드	혼합을 병렬로 나누는 머리 수		어텐션 헤드
필터 사다리	넓은 필터 몇 개부터 좁은 필터 여러 개까지 스케일이 깊이를 따라 변하는 콘볼루션 채널 변환	m_mat_w_filter	콘볼루션 필터뱅크
창	모델이 한 번에 보는 토큰 자리들. 측정 단위로는 고정 길이 텍스트 조각	T	컨텍스트 윈도우
프로브	논문 수치를 목표값과 함께 재생하는 검증 스크립트	probe 폴더	재현 스크립트
얼린 모델	학습을 멈춰 고정한 발달 단계 모델	model 폴더 npz	frozen checkpoint
홀드아웃	학습에 쓰지 않은 평가 텍스트		held-out set

용어 범례 2. 바이토큰과 게이트

용어	뜻	코드 기준	일반 용어
바이토큰	토큰 하나를 데이터면과 연산면의 직교 쌍으로 두는 구조	USE_BITOKEN	
데이터면	토큰의 내용을 담는 임베딩 면. 더하기 자리로 들어간다	m_mat_w_data_axis	토큰 임베딩
연산면	토큰의 연산 지시를 담는 임베딩 면. 곱하기 자리로만 들어간다	m_mat_w_operate_axis	
더하기 자리, 곱하기 자리	데이터면이 실리는 잔차 덧셈 위치와 연산면이 작용하는 게이트 곱셈 위치		잔차 연결, 게이팅
게이트	각 채널의 통과량을 0에서 1로 정하는 밸브. 직전 토큰의 연산면이 주입된다	m_mat_w_gate	게이팅
한 칸 이동	직전 토큰의 연산면이 다음 자리의 게이트로 주입되는 한 자리 밀림		
연산이 곧 가중치	연산면 임베딩이 라벨 없이 가중치 역할로 창발하는 현상		
일관 연산자	어떤 데이터에도 일관된 방향으로 작용하는 연산자		
방향 일관성	같은 연산이 여러 데이터를 꺾는 방향들의 평균 코사인. 연산자 창발의 지표	p_direction_consistency	
회전 연산자 게이트	연산면을 오일러 회전각으로 읽는 게이트. 본문 축약은 회전 게이트	paper7_rotation.py	
기본 연산	게이트가 한 홑 지시로 곧바로 표현할 수 있는 연산		프리미티브
깊이 배제 판별	깊이를 1블록으로 제한해 게이트만 변수로 남긴 의도된 편파 측정	제7편 프로브	
자기 되먹임	자기 출력을 입력으로 되먹이며 끝 상태를 채점하는 채점 방식		자기회귀 롤아웃

용어 범례 3. 되새김과 발달

용어	뜻	코드 기준	일반 용어
되새김	기울기 없이 임베딩을 이웃끼리 당겨 뭉치게 하는 자기조직화 학습	제2편 프로브	자기조직화
동종군	씨앗에서 코사인 최근접 이웃을 모은 뜻 가까운 개념 집합		클러스터
무게중심 당김, 감쇠	평균 쪽으로 당기고 원본 쪽으로 되돌리는 되새김의 두 힘. 균형에서 평형이 창발한다	ALPHA, 감쇠율	
월드먼저	언어 이전에 감각, 공간, 논리 축을 먼저 발달시키는 커리큘럼 순서	banya_world_data	커리큘럼 학습
발달 단계	스텝 수로 정의하고 사람 성장 이름을 붙인 체크포인트 단계		
축	임베딩 안에 방향으로 형성되는 성질의 눈금. 감각, 순서, 범주		표현 축

용어 범례 4. 어휘와 메타인지

용어	뜻	코드 기준	일반 용어
음절 원자 묶음	어휘 바닥층의 음절 단위 토큰 음절 원자를 묶어 사전에 올린 개념 토큰. 어휘 단 위로만 사용한다	AtomTokenizer	문자 단위 토큰 서브워드
캐시 사전 토큰화	원자 아래층에 묶음 사전 위층을 얹은 2층 어휘 체계	제5편 프로브	서브워드 토큰화 와 대비
확실, 모호, 모름 라우팅 재해석	자기 출력 분포로 갈리는 세 앓 상태 상태에 따라 답의 처리 경로를 배정하는 것 모호 히든을 문맥과 함께 재순전파해 확실히 끌어 올리는 절차	pmax, 엔트로피 제6편 프로브	불확실성 추정 routing
유사도 정리 추론 순회 자기 발화	데이터면 유사도로 개념이 정돈되는 축 연산면을 따라 추론을 이어 가는 축 유사도 정리와 추론 순회를 엮어 스스로 말을 잇는 목표 상태		
반야프레임	시리즈의 배경을 이루는 공리 15개 체계. 본문 주 장의 근거는 아님		

기호 범례

이 편의 수식과 본문이 함께 사용하는 기호를 정의한다. 한 식 안에서만 쓰는 국소 기호는 각 식 아래의 기호 줄에 적었다.

기호	한글 명칭	정의 및 의미	자료형
A	축 일치도	대비쌍 임베딩 차분과 축 평균 방향의 코사인 유사도를 P 전체에 평균한 값. 축마다 계산하며, 무작위 기준선 0.027 대비 5배에서 27배로 측정된다	스칼라
S	순서 정합	등급 사다리에서 다음 등급이 확률 분포 상위에 뜰 확률. 축마다 계산하며, 순서 축 33에서 67퍼센트, 범주 축 0에서 4퍼센트로 측정된다	스칼라
P	대비쌍 집합	그 축으로만 다른 낱말쌍(대비쌍)들의 집합. 축 일치도 A 를 이 집합 전체에 대해 평균한다	집합
$ P $	대비쌍 개수	집합 P 의 크기, 곧 대비쌍의 개수. 합을 평균으로 만드는 정규화 분모에 사용된다	정수
(a, b)	대비 낱말쌍	P 의 원소인 낱말쌍. a 와 b 는 그 축으로만 다른 두 낱말이다	순서쌍
E_a	낱말 임베딩	낱말 a 의 임베딩 벡터. E_b 는 낱말 b 의 임베딩으로 같은 꼴이다	벡터
$E_a - E_b$	임베딩 차분	두 낱말 임베딩의 차분으로, 그 축의 방향을 나타내는 벡터	벡터
$\bar{v}$	축 평균 방향	그 축의 평균 방향 벡터. 대비쌍 차분들이 이 방향과 일치하는 정도로 축 형성을 읽는다	벡터
$\cos$	코사인 유사도	두 벡터 사이 각도의 코사인으로 방향 일치 정도를 측정하는 함수. 4.3절의 중심 코사인도 같은 척도이다	함수
$\sum_{(a,b) \in P}$	합 기호	P 의 모든 대비쌍에 걸쳐 코사인 유사도를 더하는 합 연산자	연산자
$\Pr[\cdot]$	확률	대괄호 안 사건이 일어날 확률. 순서 정합 S 의 정의에 사용된다	함수

1. 서론

사람이 직접 개념 지도를 짜 넣는 온톨로지 구축은 오래 걸리고 늘 빈틈이 있다. 본고는 반대로 접근한다. 온톨로지를 짜 넣지 않고, 학습 순서를 사람의 발달 순서에 맞춘다. 언어 이전에 감각과 공간과 논리 축을 먼저 발달시키면, 필터가 그 축들을 스스로 형성해 온톨로지 역할을 한다는 가설이다. 이 관점의 형식적 배경은 별도 공리 문헌 [1]에 있으나, 본고는 그 가설이 실제 학습에서 성립하는지를 측정으로만 검증한다.

학습에는 오차역전파를 자동미분 없이 계산하는 제1편 [2]의 엔진을 쓴다. 판독은 학습을 멈추지 않는 비침습 프로브로 한다. 축 대비쌍(그 축으로만 다른 낱말쌍)의 임베딩 차분 방향이 서로 얼마나 일치하는지를 무작위 기준선과 비교해 축이 형성됐는지 읽고, 등급 사다리를 넣어 다음 등급이 분포에 뜨는지로 순서가 확립됐는지 읽는다.

2. 관련 연구

커리큘럼 학습(curriculum learning) [5]은 쉬운 예에서 어려운 예로 학습 순서를 잡지만, 대개 난이도 정렬이며 감각 축을 언어 앞에 형성하지는 않는다. 소데이터 발달 학습을 거르는 시도들도 언어 코퍼스 자체의 순서를 다루지, 언어 이전의 세계 축을 먼저 형성하지 않는다. 기계적 해석성(mechanistic interpretability, 신경망 내부가 무엇을 표현하는지 뜯어보는 연구)은 표현을 사후에 읽는다. 본고는 학습

순서(월드먼저)를 독립 변수로 두고, 그 결과로 창발한 축 구조를 비침습으로 판독한다는 점이 다르다.

3. 방법

발달 커리큘럼은 언어 이전 단계(월드)에서 감각, 공간, 논리 축을 먼저 학습하고, 이어 유아, 유아2 단계의 언어를 엮는다. 각 단계는 이름 붙은 체크포인트(탄생, 아기, 유아 등)로 열려 두어 단계별로 축 형성을 측정한다.

식 3-1. 축 일치도와 순서 정합

$$A(\text{축}) = \frac{1}{|P|} \sum_{(a,b) \in P} \cos(E_a - E_b, \bar{v}), \quad S(\text{축}) = \Pr[\text{다음 등급이 확률 상위에 분포}]$$

수식

기호: P 는 그 축으로만 다른 낱말쌍의 집합, $E_a - E_b$ 는 두 낱말 임베딩의 차분(그 축 방향), $\bar{v}$ 는 그 축의 평균 방향, $\cos$ 은 코사인 유사도, A 는 축 일치도, S 는 순서 정합이다.

작동 방식: 같은 축의 낱말쌍들이 임베딩에서 같은 방향으로 놓이면 A 가 크다. 무작위 기준선보다 몇 배 크면 그 축이 임베딩에 실재한다. S 는 등급 사다리에서 다음 등급이 분포 위에 뜨는 확률로, 그 축이 순서로 정렬됐는지를 측정한다.

실제 동작: 예로 크기 축은 "작다, 크다" 같은 등급쌍으로 측정한다. 이 쌍들의 차분 방향이 일치하면 크기 축이 형성됐고, "작다" 다음에 "조금 크다"가 분포 위에 뜨면 순서로 정렬된 것이다.

4. 결과

4.1 감각 축이 무작위의 5배에서 27배로 형성된다

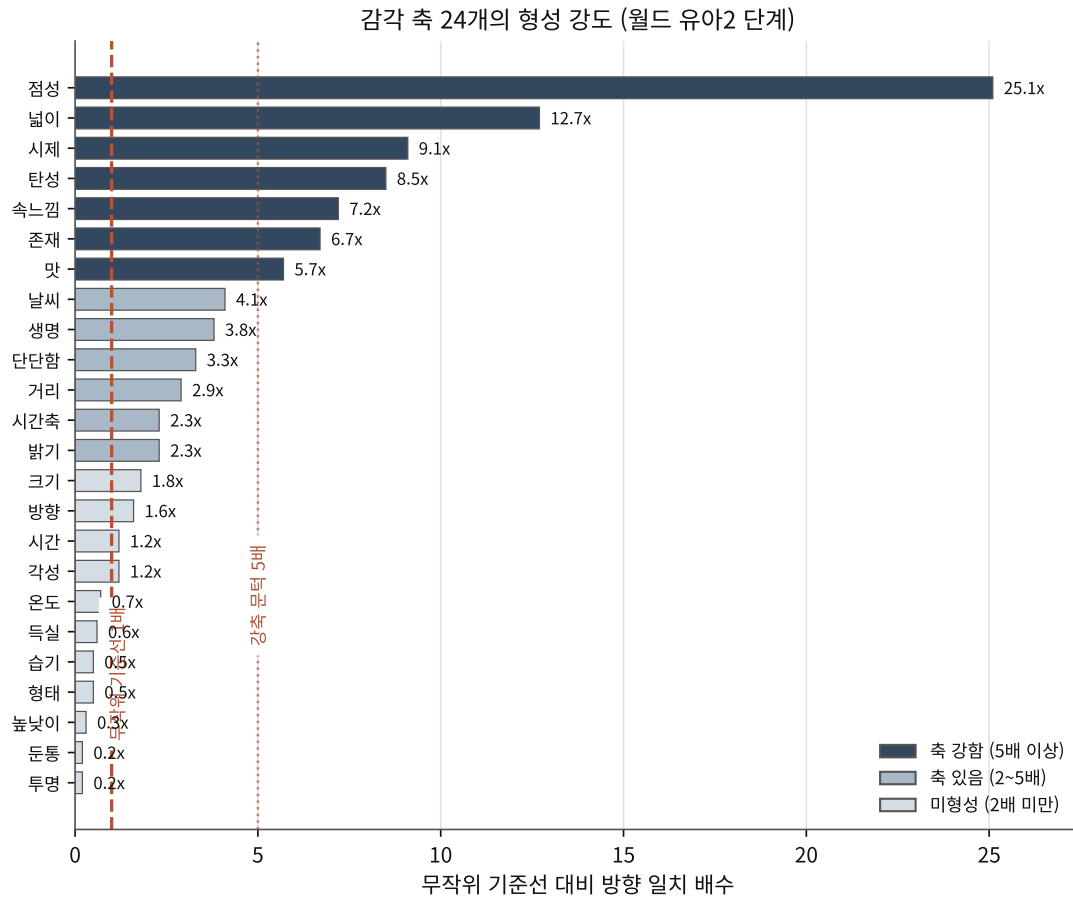
측정

발달 단계 열린 모델에서 축 일치도를 측정한 결과(무작위 기준선 0.027), 여러 감각 축이 강하게 형성된다. 탄생부터 유아 단계까지 안정적으로 강한 축의 기준선 대비 배수:

축	기준선 대비	축	기준선 대비
점성	24 ~ 27배	시제	8 ~ 10배
넓이	11 ~ 13배	존재	7배
날씨	4 ~ 11배	맛	6배
탄성	4 ~ 9배	속느낌	5 ~ 7배

단계마다 강한 축이 7개에서 8개로 안정되게 유지된다. 모든 축이 형성되는 것은 아니며(일부 축은 미형성), 균일하게 세지는 것도 아니지만, 감각 개념이 무작위보다 현저히 강하게 축에 자리한다는 사실은 모든 단계에서 일관된다.

그림 4-1. 감각 축 일치도 (무작위 대비 배수)



여러 감각 축이 무작위 기준선의 5배에서 27배로 형성된다. 개념을 직접 짜 넣지 않아도 축에 자리한다. 막대 길이는 무작위 대비 배수이며 세로선 1배가 무작위 수준이다.

4.2 순서 축은 순서로, 범주 축은 순서 없이 형성된다

축에는 순서가 있는 것(거리, 온도)과 순서가 없는 범주(색, 맛)가 있다. 측정 결과 이 둘이 명확히 갈린다.

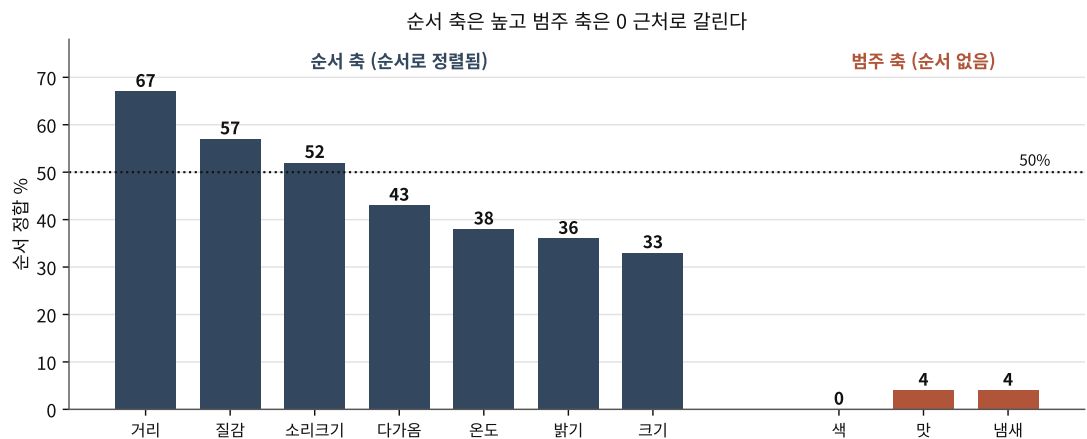
측정

순서 정합(다음 등급이 분포 위에 뜨는 비율) 측정값:

- 순서 축 : 거리 67퍼센트, 질감 57퍼센트, 소리크기 52퍼센트, 다가옴 43퍼센트, 온도 38퍼센트, 밝기 36퍼센트, 크기 33퍼센트.
- 범주 축 : 색 0퍼센트, 맛 4퍼센트, 냄새 4퍼센트.

순서 축은 33에서 67퍼센트로 순서 정합이 높고, 범주 축은 0에서 4퍼센트로 순서가 없다. 모델이 순서 있는 성질은 순서로, 종류만 있는 성질은 순서 없이 배치했다. 이 갈림 자체가 축 구조가 창발했다는 증거이다.

그림 4-2. 순서 축과 범주 축의 순서 정합



순서가 있는 축(거리 67퍼센트 등)은 순서로 정렬되고, 순서가 없는 범주 축(색 0퍼센트 등)은 순서 없이 형성된다. 모델이 성질의 종류에 맞게 배치했다. 이 명확한 갈림이 축 구조 창발의 직접 증거이다.

4.3 관련 개념끼리 무리를 이룬다

축 사이의 관계도 창발한다. 측정 결과 공간 개념끼리(거리와 다가옴의 중심 코사인 0.84, 방향과 거리 0.63), 감각 개념끼리(냄새와 각성 0.94, 맛과 냄새 0.93) 무리를 이룬다. 관련 있는 축이 임베딩에서 서로 가까이 모여, 온톨로지의 상위 묶음처럼 자리한다.

4.4 채널 필터가 필터 사다리로 자기조직하며 죽은 필터가 없다

측정

채널 필터를 측정한 결과, 얇은 층에서 깊은 층으로 갈수록 필터가 넓고 적은 것(큰 붓)에서 좁고 많은 것(가는 펜)으로 필터 사다리를 이룬다(길이 1024의 1개에서 길이 32의 32개까지). 이 구조는 제3편 [4]의 필터 사다리다. 모든 층에서 죽은 필터(응답이 약해 사실상 안 쓰이는 필터) 비율이 0퍼센트이다. 게이트의 입력 의존 채널 비율은 얇은 층 94퍼센트에서 깊은 층 61퍼센트로, 대부분의 채널이 입력에 따라 반응한다.

그림 4-3. 채널 필터의 필터 사다리와 생존

얇은 층(큰 붓) ▶ 깊은 층(가는 펜), 죽은 필터 0%



깊이가 깊어질수록 필터가 좁고 많아진다. 어느 층에도 죽은 필터가 없다(생존 100%).

채널 필터가 얇은 층 큰 붓에서 깊은 층 가는 펜으로 필터 사다리를 이루며 자기조직한다. 모든 층에서 죽은 필터가 0퍼센트로, 발달 학습이 필터 용량을 낭비 없이 쓴다.

재현

4절의 축 구조 수치는 부록 C의 준비를 마친 뒤 다음 한 줄로 재생된다.

```
cd probe && python3 paper4_world_dev_probe.py
```

출력에서 감각 축 배수, 순서 축과 범주 축의 정합, 필터 사다리와 생존율이 본문 수치와 대응한다. 이 프로브는 말뚝치가 필요 없어 부록 C의 4단계를 건너뛰어도 된다.

5. 한계와 다음 측정 과제

기준선은 1인 개발, 자작 정확한 전치 역전파 엔진, 소비자 그래픽 처리장치 한 대라는 레퍼런스 클래스이다. 아래는 결함이 아니라 다음 측정 과제이다.

- 순서 대조 : 같은 데이터와 스텝에서 학습 순서만 뒤집은 언어먼저 대조를 나란히 놓아 순서 효과를 격리하는 것이 다음 과제이다. 순서가 원인임을 못 박으려면 이 대조가 필요하다.
- 후기 단계 능력 : 질의응답이나 자아 표현 같은 상위 능력은 본고가 측정한 유아 단계 열린 모델에서는 아직 낮게 나온다. 이는 더 뒤 단계의 몫이며 제6편에서 다룬다. 본고는 그 능력을 주장하지 않는다.
- 표준 해석성 대조 : 자작 프로브 외에 표준 해석성 기법과의 나란한 비교가 다음 과제이다.

6. 결론

본고는 온톨로지를 직접 짜 넣지 않고 학습 순서를 사람의 발달 순서에 맞추면, 필터와 임베딩이 스스로 축 구조를 형성함을 측정으로 보였다. 감각 축이 무작위의 5배에서 27배로 형성되고, 순서 축은 순서로 범주 축은 순서 없이 갈리며, 관련 개념끼리 무리를 이루고, 채널 필터가 필터 사다리로 자기조직하되 죽은 필터가 없다. 학습 순서 자체가 필터를 온톨로지로 만든다는 가설이, 측정한 단계에서 축 구조의 창발로 확인된다. 상위 능력의 확인과 순서 대조는 다음 과제로 둔다. 말보다 세계를 먼저 넣는 이 순서는 본고가 처음 그린 것이 아니라, 반야프레임의 공리가 가리키는 더 오래된 밑그림에서 왔다.

7. 후속 논문 예고

후속 논문 예고

- 제5편 — 캐시 사전 토큰화와 동적 임베딩 : 본편의 축이 얹히는 어휘층을 음절 원자와 묶음 사전의 2층 구조로 다룬다.
- 제6편 — 자기 분포와 기하 모니터링에서 창발하는 제어 신호 : 본편이 후기 단계로 미룬 상위 능력(확실, 모호, 모름의 라우팅, 학습 동기)을 측정된 통계량으로 다룬다.

- 마지막 편(미완) — 유사도 정리와 추론 순회의 결합 : 데이터면 유사도 정리(제2편 [3])와 연산면 추론 순회를 엮는 자기 발화. 최종 조립이 끝나지 않아 아직 다루지 않는다.

참고문헌

- [1] 한혁진 (2026). Banya Framework: An Axiom-Based Science Mining Engine for Deriving Fundamental Physical Constants (반야프레임 공리 15개). Zenodo. <https://doi.org/10.5281/zenodo.20301304> . (본 시리즈의 배경 자연관, 선행 공리 문헌)
- [2] 한혁진 (2026). 자동미분 없이 수동으로 유도한 정확한 전치로 학습하는 언어모델 엔진. 반야 연구 시리즈 제1편. Zenodo. <https://doi.org/10.5281/zenodo.21385447>
- [3] 한혁진 (2026). 기울기를 전혀 쓰지 않는 되새김 자기조직화 학습. 반야 연구 시리즈 제2편. Zenodo. <https://doi.org/10.5281/zenodo.21385516>
- [4] 한혁진 (2026). 바이토큰: 직교 데이터-연산 토큰과 라벨 없이 창발하는 연산자. 반야 연구 시리즈 제3편. Zenodo. <https://doi.org/10.5281/zenodo.21385676>
- [5] Bengio, Y., Louradour, J., Collobert, R., Weston, J. (2009). Curriculum Learning. ICML.

부록 A. 발달 단계

표 A-1. 측정에 쓴 발달 단계 열린 모델

단계	스텝	단계	스텝
탄생	1만	유아	7만
아기	3만	유아2	11만

부록 B. 측정 환경

표 B-1. 측정에 사용한 컴퓨터 사양

구분	사양
그래픽 처리장치(GPU)	NVIDIA GeForce RTX 3090 Ti (24 GB)
운영체제	Ubuntu 24.04.4 LTS
소프트웨어	Python 3.12.3, numpy 2.5.1, scipy 1.18.0, cupy 14.1.1, CUDA 13.3.1
측정 대상	발달 단계별 열린 모델의 임베딩, 축 대비쌍, 채널 필터

4절의 실측 수치는 위 환경에서 측정했다.

부록 C. 재현 안내

본고의 실측 수치는 공개 재현 패키지로 재생된다. 아래 다섯 단계를 순서대로 하면 된다.

1. 요구 환경. NVIDIA 그래픽 처리장치(GPU)와 CUDA, Python 3.12 가 필요하다. 파이썬 패키지는 압축을 푼 폴더에서 다음으로 설치한다. cupy 는 설치된 CUDA 버전에 맞는 배포판을 사용한다(예: CUDA 13 은 cupy-cuda13x).

```
pip install -r requirements.txt
```

2. 코드 받기. 압축 묶음(zip)을 내려받아 푼다. 저장소로 받으려면 github.com/ubmscoin/banya/banya_ai_paper_code 폴더다.

```
wget https://ubmscoin.github.io/banya/banya_ai_paper_code/
banya_ai_paper_code.zip
unzip banya_ai_paper_code.zip && cd banya_ai_paper_code
```

3. 모델 받기. 열린 모델 3개(합계 471MB)는 제노도 doi.org/10.5281/zenodo.21383724 에 있다. 다음 한 줄이 내려받기와 무결성 검증(sha256, 파일 지문 대조)을 함께 한다. 손으로 받으려면 위 기록에서 세 파일을 받아 model 폴더에 넣고 sha256sum -c checksums.sha256 으로 검증한다.

```
cd model && bash download.sh && cd ..
```

4. 말뭉치 만들기. 말뭉치는 원문 배포 없이 생성기가 만든다. 다음 한 줄이 banya_world_data 폴더에

말뭉치 23종을 전부 생성한다.

```
cd corpus_build && bash build_all.sh && cd ..
```

5. 실측 재생. probe 폴더의 프로브가 각 절의 수치를 재생한다. 본고의 수치는 다음 프로브가 담당한다. 출력에 목표값이 함께 인쇄되므로 본문과 바로 대조된다.

프로브	재생하는 수치	확인할 출력
paper4_world_dev_probe.py	4절 축 형성과 필터 사다리	축 배수와 순서 정합, 필터 사다리 수치가 목표값과 함께 인쇄된다. 말뭉치 불필요

패키지의 폴더 구성과 파일별 설명은 [목차 페이지](#)에 있다.

반야 연구 시리즈 제4편(월드먼저 발달, 창발 축 구조). 모든 수치는 부록 B 환경에서 측정한 값이다. 배경 자연관은 선행 공리 문헌 [1]로 인용하되 이 글 주장의 근거가 아니다. 제1편에서 제3편을 인용한다. 질의응답과 자아 같은 후기 단계 능력은 측정 단계에서 아직 낮으며 제6편에서 다룬다. 캐시 사전 토큰화는 제5편에서 다룬다.